

---

**A Relational Fairness  
Perspective on Individuality  
and Individualization  
in Contemporary Approaches to  
Compensation in Organizations**

---



**A Relational Fairness  
Perspective on Individuality and  
Individualization in Contemporary  
Approaches to Compensation in  
Organizations**  
*A Study of Social Rationality, Difference,  
and Differentiation*

---

*Revised Version*

**PhD Thesis  
Defended in Public at  
University of Southern Denmark  
(Spring 2014)**

**Martin Lund Petersen**

*Department of Leadership and Corporate Strategy  
University of Southern Denmark  
Slagelse, Denmark*

  
**River Publishers**

ISBN 978-87-93237-30-8 (e-book)

*Published, sold and distributed by:*

River Publishers  
Niels Jernes Vej 10  
9220 Aalborg Ø  
Denmark

Tel.: +45369953197  
[www.riverpublishers.com](http://www.riverpublishers.com)

Copyright for this work belongs to the author, River Publishers have the sole right to distribute this work commercially.

All rights reserved © 2014 Martin Lund Petersen.

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without prior written permission from the Publisher.

# Table of Contents

English-Summary.....	5
Dansk Resume.....	8
Foreword.....	11
Problem Statement.....	18
Chapter I.....	20
1.1 Introduction.....	20
1.2 Knowledge and Theoretical Construction.....	20
1.2.1 Causal Mechanisms and the Dimension of the Real.....	23
1.3 Theoretical Delimitations.....	38
Chapter II.....	43
2.1. Introduction.....	43
2.2 Motivating the Individual to Perform.....	46
2.3 Individualization.....	50
2.3.1 Pay the Individual.....	50
2.3.2 Individualization of the Employment Relation.....	56
2.4 Recognizing the Individual.....	58
2.4 Individual Differentiation, Social Norms, and the Need to Belong.....	62
2.5 Tentative Conclusions.....	68
Chapter III.....	71
3.1 Introduction.....	71
3.2 Introduction to Fairness Theories.....	71
3.2 On Rationality and Morality.....	81
3.2.1 Plato.....	81
3.3.2 Thomas Hobbes.....	100
3.3.3 Adam Smith.....	122
3.3.4 David Hume.....	133
3.3.5 Immanuel Kant.....	138
3.3.6 Jeremy Bentham and Economic Rationality.....	163
3.3.7 Jon Elster and Donald Davidson.....	171

3.3.8 Raymond Boudon and the Positive and Normative Dimension of Rationality .....	180
3.4 Tentative Conclusions .....	187
Chapter IV.....	198
4.1 Introduction.....	198
4.2 Economic Rationality and Fairness as Reciprocity .....	200
4.3 On Authority and Renegotiation .....	217
4.4 Authority, Incentives, and Non-economic Motivation.....	220
4.5 Tentative Conclusions .....	221
Chapter V.....	225
5.1 Introduction.....	225
5.2 The Balance, the Middle, and Cognitive Dissonance .....	229
5.3 Social Utility and Social Preferences .....	238
5.4 Tentative Conclusions .....	246
Chapter VI.....	248
6.1 Introduction.....	248
6.2 Cognitive Normative Expectations and Relative Deprivation .....	249
6.3 The Foundation of Cognitive Normative Expectations .....	251
6.4 Fairness and Blame .....	253
6.5 Tentative Conclusions .....	261
Chapter VII.....	263
7.1 Introduction.....	263
7.2 Morality as Irrationality.....	264
7.3 Hedonic Experience as the Foundation of Fairness .....	269
7.4 On the Foundations of Expectations.....	276
7.5 Tentative Conclusions .....	282
Chapter VIII.....	287
8.1 Introduction.....	287
8.2 Social Norms.....	288
8.3 The Need to Belong and Social Norms.....	299
8.3.1 In Search of an Understanding of the Pragmatic Necessity of Relatedness .....	305
8.5 Tentative Conclusions .....	312
Chapter IX.....	316
9.1 Introduction.....	316

9.2 Research Design .....	317
9.3 Asking Sensitive Questions with Difficult Answers .....	322
9.4 Vignettes .....	324
9.5 Mechanisms and Analytical Strategy .....	326
9.6 The Qualitative Study .....	329
9.6.1 Interviews .....	330
9.6.2 Epistemological Considerations .....	334
9.6.3 Interview-Guide .....	335
9.6.4 Respondent Group .....	338
9.6.5 Transcription of Interviews .....	340
9.6.6 Quality Issues Related with the Qualitative Design .....	340
9.7 The Quantitative Study .....	342
9.7.1 Survey .....	343
9.7.2 Measuring-Scale .....	347
9.7.3 Epistemological Considerations .....	348
9.7.4 Respondent Group .....	349
9.7.5 Quality Issues Related with the Quantitative Design .....	352
9.8 Methodological Delimitations .....	353
Chapter X .....	356
10.1 Introduction .....	356
10. 2 Presentation of the cases .....	356
10.3 Variables .....	357
10.3.1 Independent Variables .....	358
10.3.1 Dependent variables .....	359
10.4 Two Basic Scenarios .....	360
10.4.1 Exogenous Differences .....	361
10.4.2 Endogenous Differentiation .....	361
10.5 Analysis and Results .....	362
10.5.1 Exploring the Cases .....	363
10.5.2 Fairness and Norm Violation .....	385
10.5.3 Fairness, Norm Violation, and Emotions .....	388
10.5.4 Relational Influence .....	394
10.6 Tentative Conclusions .....	403

Conclusion .....	410
References.....	430
Appendix-1: Translated-Edition of the questionnaire.....	450
Appendix-2: Cover Letters.....	453
Appendix-3: Interview Summaries.....	458
Appendix-4: DVD with Interviews	

## English-Summary

The primary concern of this thesis is to explore the relations between contemporary compensation systems based on individualization and individuality, the constitution of the fair claim, the causality of blame and the causality of rewarding, and the need to belong and its relation with human motivation. More specifically, the primary research question on which this thesis is based is: How is the fair claim constituted in, respectively, a compensation system based on individualization and individuality, a social relational perspective; and in which ways might potential differences influence horizontal relations?

The philosophical foundation of this thesis is critical realism. This perspective implies a focus on the causal mechanisms which produce or constitute the fair claim – hence, the focus is not on how these claims manifest themselves empirically, rather the focus is on their normative relational constitution. From this perspective also follows a particular approach to the research process, insofar as this must, to some extent, be able to capture these generative mechanisms, thus the research process is based on the combination of theoretical abstractions and empirical data. The empirical foundation of this thesis is a small survey and group/personal interviews. The survey has as its aim to collect data on across respondent tendencies in fairness judgments, whereas the interview study has as its aim to collect data on the cognitive or associative pattern on which these fairness judgments are centered.

In chapter II, I discuss and analyze how the fair claim is constituted in a compensation system based on individuality and individualization. The fair claim produced by such a system cannot be reduced to matters solely related with obtaining a fair share, insofar as the values are more general in nature. Individuality is related with designing the compensation system in a manner which allows it to recognize and reward individuality – which implies a focus on individual performance, just as it implies that the employee must be allowed to design the reward mix in accordance with her interests. Individuality might, then, lead to compensatory differences in outcome related with contingent rewards and base-pay based on difference in competency, just as different choices

might lead to differentiation and individualization of the employment relation. The fair claim may, in other words, be based on recognizing individuality. Horizontally, the claim may be defensible because it reflects individuality and is based on accurate measures – thus it is rationally acceptable. The questions might, however, be: 1) whether vertical relations can be conceptualized without some moral or social normative foundation, 2) whether horizontal fairness can be reduced to rational acceptance, and 3) whether a compensation system unaligned with the social and moral normative relational foundation might undermine these and subsequently due to the social psychologically need to belong undermine the motivation it is designed to bring about.

In chapter III, I discuss and analyze the relation between rationality and morality in order to understand whether morality is irrationality, whether rationality is actually reducible to instrumental or economic rationality. Firstly, I discuss the metatheoretical foundation of theories on organizational fairness. Secondly, I identify the intellectual historic foundation of these theories on organizational fairness by discussing and analyzing different philosophical approaches to the relation between morality and rationality. In chapter IV, I discuss how the vertical relation might be based on a normative foundation similar to the quasi-moral principle of reciprocity, just as I also argue that while this might provide some insight into the normative foundation, one should not neglect the formalized nature of employment relation.

Whereas the relation between morality and rationality may provide some insight into the governance of social interaction by identifying the moral and social bounds which binds individuals to one another, it might not provide any insight into why outcome differences are accepted. Hence, in chapters V-VII, I discuss and analyze how the fair claim may be constituted on three normative dimensions; relational and temporal status quo, and cognitive normative expectations. Finally, in chapter VIII, I discuss the relation between the rational, social, and moral foundation of interaction, the social psychological need to belong, and human motivation to achieve or excel.

Using this theoretical frame as point of departure, I study how the fair claim from a third-person perspective is constituted empirically in fairness judgments. The study shows, among other things, that in some situations the normative foundation may not be as important as other legitimate claims, e.g. achievement. Additionally, my study also

shows that in situations considered as horizontally as well as vertically fair, a potential outcome difference may have no relational influence.

In conclusion, then, what is argued here is not that these contemporary compensation systems cannot function, rather the argument is that one has to take into consideration the social and moral normative foundation, insofar as a difference perceived as vertically and horizontally unfair might have a negative relational influence which might through the need to belong lead to an undermining of the motivation to excel and to individuality.

## Dansk Resume

Det primære fokus i nærværende afhandling er at belyse relationerne mellem moderne kompensationsystemer baseret på individualisering og individualitet, konstitutionen af den retfærdige ret, skylds kausalitet og belønnings kausalitet, og behovet for sociale relationer og dets relation til motivation. Mere specifikt er hovedspørgsmålet for denne afhandling: Hvordan er den retfærdige ret konstitueret i henholdsvis et kompensations-system baseret på individualitet og individualisering, et socialt relationelt perspektiv og hvordan vil eventuelle forskelle påvirke horisontale relationer?

Det filosofiske fundament for afhandlingen er kritisk realisme. Dette perspektiv omfatter bl.a. et fokus på de kausale mekanismer, der producerer eller konstituerer den retfærdige ret og således ikke kun, hvordan denne manifesterer sig empirisk. Fokus er derfor på rettens normative relationelle konstitution. I forlængelse af dette perspektiv følger også en særlig tilgang til selve forskningsprocessen i den forstand, at denne skal muliggøre indfangelsen af disse generative mekanismer. Forskningsprocessen er derfor baseret på en kombination af teoretiske abstraktioner og empirisk data. Den empiriske undersøgelse er baseret på en mindre spørgeskemaundersøgelse or gruppe/personlige interviewundersøgelse. Spørgeskemaundersøgelsen har som sit mål at indsamle data om tendenser i retfærdighedsdomme på tværs af respondenterne, hvorimod interviewundersøgelsen har som sit mål at indsamle data om de kognitive strukturer som disse domme er baseret på.

I kapitel II diskuterer og analyserer jeg, hvordan den retfærdige ret er konstitueret i et kompensationsystem baseret på individualitet og individualisering. Den retfærdige ret produceret af sådanne systemer kan dog reduceres til udelukkende at omhandle opnåelsen af et retfærdigt udbytte, idet værdierne er mere grundlæggende i deres natur. Individualitet er relateret til, at kompensationsystemet er designet på en måde, der muliggør anerkendelse og belønning af individualitet. Dette medfører bl.a. et fokus på individuel performance ligesom det også medfører, at medarbejderen har lov til at designe sin belønningspakke efter egne interesser. Individualitet kan derfor føre til forskelle i udkomme baseret på adfærdsafhængig belønning og en basisløn baseret på kompetence forskelle, ligesom forskellige valg også kan lede til differentiering og individualisering

af medarbejderrelationen. Den retfærdige ret er m.a.o. baseret på anerkendelsen af individualitet. Horisontalt er retten retfærdiggjort fordi den afspejler individualitet og er baseret på nøjagtige mål, hvorfor den er rationelt acceptabel. Der kan dog stilles spørgsmål ved: 1) om vertikale relationer kan begribes uden et moralsk og/eller socialt normativt fundament, 2) om horisontal retfærdighed kan reduceres til et spørgsmål om rationel accept, og 3) om et kompensationsystem som ikke er kompatibelt med det sociale og moralske normative relationelle fundament ikke vil underminere disse og i sidste ende, som en konsekvens af det sociale tilhørsbehov underminere den motivation som det var designet til at skabe.

I kapitel III diskuterer og analyserer jeg relationen mellem rationalitet og moralitet med henblik på at belyse, hvorvidt moralitet kan sidestilles med irrationalitet og hvorvidt det er muligt at reducere rationalitet til instrumentel eller økonomisk rationalitet. Først diskuterer jeg det metateoretiske fundament for teorier om organisatorisk retfærdighed og derefter identificeres det videns genealogiske fundament for disse ved at diskutere forskellige filosofiske tilgange til relationen mellem rationalitet og moralitet. I kapitel IV diskuterer jeg, hvordan vertikale relationer kan begribes som værende centret på det kvasi-moralske princip, reciprocitet. I forlængelse heraf argumenterer jeg dog også for at dette perspektiv kun omfatter den normative dimension og derfor ikke hele kompleksiteten som følger fra denne relations særlige natur.

Hvor relationen mellem rationalitet og moralitet kan belyse de bagvedliggende strukturer for social interaktion ved at identificere de moralske og sociale bånd, der binder individerne til hverandre er dennes blinde plet måske, at den ikke giver indsigt i, hvorfor nogle forskelle er accepteret. I kapitlerne V-VII diskuterer og analyserer jeg derfor, hvordan den retfærdige ret kan begribes som værende baseret på tre normative dimensioner: et relationelt og temporalt status quo, og kognitivt normative forventninger. Endelig i kapitel VIII diskuterer jeg relationen mellem det rationelle, sociale og moralske fundament for interaktion, det psykologiske behov for socialtilhør og motivationen til selv-realiserings.

Med udgangspunkt i denne teoretiske ramme undersøger jeg, hvordan den retfærdige ret fra et tredjepersons perspektiv konstitueres i empiriske domme om retfærdighed. Min undersøgelse viser bl.a., at det normative fundament i nogle situationer kan tilside-

sættes af andre legitime begrundelser fx indsat. Min undersøgelse viser også at i situationer opfattet som horisontalt og vertikalt retfærdige vil eventuelle belønningsforskelle ikke have en negativ horisontal relationel indflydelse.

En samlet konklusion er derfor ikke, at sådanne moderne kompensationssystemer ikke fungerer, men derimod at det er nødvendigt at tage det sociale og moralske normative fundament i betragtning, idet en eventuel horisontal og vertikal uretfærdig forskel kan have en negativ relationel indflydelse som kan føre til en forringelse af det psykologiske behov for socialtilhør som i sidste ende kan underminere lysten til selvrealisering – lysten til individualitet.

## Foreword

As the title of this thesis suggests, the present work is an analysis and discussion of the relational fairness of contemporary compensation strategies in organizations and in particular the two underlying values, individuality and individualization. The subtitle reveals another dimension of this work by referring to the social manifestations of relational fairness, individuality, and individualization in social rationality, compensatory differences, and self-differentiation. *Relational* refers to the fact that organizations are vertically as well as horizontally structured and characterized by three basic relations; organization-employee, manager-employee, and employee-colleagues. *Fairness* refers to a particular type of evaluative judgments which have as their area of concern evaluations of whether something is just, alright, acceptable, in accordance with governing normative expectations, etc. *Individuality* refers to a particular trend in contemporary strategies on compensation related with a focus on individuality and a growing concern with designing compensation strategies in a manner which respects the employee's individuality – recognition of the fact that employees are different and that they are responsible rational agents who should be allowed to let their individuality be reflected in different choices, while *individualization* refers to the results of this focus on respecting individuality, manifested in, for example, the individualization of the employment relation.

Recognition of individual differences and a growing focus on recognizing not only individuality, but also individual performance differences can lead to compensatory differences simply just between different groups of employees, but also differences within these groups. Whereas the concept of difference captures outcome differences, the concept of differentiation captures the acts of intragroup group differentiation which brings about these states of difference – that is, qualitatively it captures the results of this recognition of individuality related with allowing the employees to design in accordance with their preferences the reward mix and quantitatively it captures differences in, for example, effort. *Social rationality* refers to how these relational fairness concerns manifests themselves in interactions – vertically (organization-employee and manager-employee) as well as horizontally (employee-organization). As such, social rationality is

concerned with social behavior – that is, how different reward practices in vertical relations takes on particular symbolic meanings and how these acts of intragroup differentiation may be governed by different horizontal normative expectations.

From this it seems to follow that matters related with fairness are caused in the sense that they are brought about by actions originating from the vertical and horizontal relations. That is, the effects of a compensation system manifest itself in vertical and horizontal practices. Fairness is related with actions as well as their consequences. Firstly, fairness cannot be reduced to a consequentialistic perspective, insofar as it has to take into account actions; secondly, the focus on actions implies a non-consequentialistic focus on the internal causal process which precedes the action and its potential consequences; and thirdly, extending beyond the distinction consequentialism/non-consequentialism, this focus on fairness/unfairness as states brought about, implies a causal focus and by taking into consideration the two basic relations, it also implies that one has to distinguish between causality as related with consequences stemming from vertical relations and causality as related with consequences stemming from horizontal relations, insofar as one might be able to establish a direct causal relationship between actions performed by management and their consequences evaluated as fair/unfair, whereas the consequences of actions originating from horizontal relations, for example, acts of differentiation and their consequences in terms of bringing about compensatory differences cannot be directly established, rather it implies the mediation of management, because the employee differentiating herself from her colleagues does not decide whether her efforts are rewarded or not.

The main concern of this thesis is, in other words, that contemporary compensation strategies focusing on rewarding the individual and recognizing individuality by allowing choice and contingent rewarding also necessitates a focus on the causal processes of compensation systems without, however, reducing this solely to a focus on vertical fairness, insofar as the focus on individual actions and contingent rewards also implies a horizontal dimension following simply from the fact that if individuality and individualization is to be rewarded and recognized, it must be based on the employees' choices to invoke these as principles of action. Additionally, the focus on the individual and not the group, redirects the focus of compensatory differentiation from intergroup to

intragroup, just as it implies a shift in causality from a perspective on differentiation as something which is done to the group in terms of external causality to something originating from within the group in terms of internal causality – hence, the classic distinction between the work-group and management cannot be maintained, but must be substituted with the horizontal relation. That is, the contemporary strategies in compensation management necessitate the introduction of the horizontal relation into the causality of compensation.

This particular focus on the vertical and horizontal structuring of organizations and their implications for compensation systems was originally inspired by Jon Elster's (1989: 215) discussion of labor-labor and labor-capital bargaining in wage determination. However, in juxtaposition to Elster, I do not solely focus on conflicting interests emerging from different normative systems generating different normative sets of expectations, rather I try to take seriously the different roles played by these relations in the causal processes of compensation, just as my main focus is not the classic dichotomy between management or capital and workers, insofar as understanding fairness judgments in contemporary compensation systems implies a different focus and a more thorough understanding of causality and the cognitive foundation of judgments on fairness/unfairness. Notwithstanding this, Elster's approach to the phenomenon inspired my initial concern and has guided my studies into the *mechanics* of compensation processes. Besides being inspired by Elster's approach, the present approach is also heavily inspired by Edward Lawler III's approach to compensation and in particular his 1990 work on strategic compensation, which inspired the *New Pay* perspective in contemporary theories to compensation management. The aim of this thesis is not so much to develop a critique of Lawler's theory as it is to understand the implications of his theory in terms of fairness, insofar as the main argument in this thesis is that if fairness is to be understood within this new approach, one has to go beyond an isolated focus on vertical fairness and take into consideration the implications of horizontal relations.

A focus on the functional implications of different norm sets and the their conflicts is, as such, not new, but was identified in the Hawthorne Studies and is often associated with some of key findings from the *Bank Wiring Room* study. Nevertheless, the main concern here is not whether functional implications can be identified, rather the

focus is on the formation of fairness judgments and the fact that these cannot be reduced to vertical relations, that is, a compensation system can be evaluated as being fair or unfair based on its consequences (distributive fairness), its procedures (procedural fairness), or how it is implemented in actual practices (interactional fairness); but the focus on individuality and contingent rewards also necessitates that the employees enact these principles – in the sense that if, for example, individual high performance is to be rewarded an employee must differentiate herself from her colleagues in order to stand out as a high performer and be rewarded as such, that is, of course, if the internal causality notion implied in compensation strategies is to be taken seriously. Hence, from a vertical perspective and the perspective of the receiving agent, the reward she obtains may be fair or unfair, but the actions she performed in order to bring this situation about might from a horizontal perspective be viewed as unfair – as if she violated some moral or social norm in order to obtain this reward. Thus, it might be questioned whether a compensation system based on delivering fair consequences is also horizontally fair if the actions needed to bring these vertically fair rewards about are unfair? Studies like, for example, the Hawthorne seem to suggest that they are not. However, instead of focusing on different norm set, the question asked in this thesis is more basic in nature, in the sense that it seeks to identify the constitution of fairness claims horizontally as well as vertically, because it is centered upon the assumption that compensation practices unfolds themselves within a social space and, as such, they must be embedded in the complex exchanges of rights and obligations on which social life is centered – that is, the question becomes, why some compensatory differences accepted while others are not and from a horizontal perspective this perspective necessitates a focus on the actions performed by the receiving and non-receiving individuals.

From a somewhat different perspective, one might also argue that whereas the crucial concern in compensation strategies from the 1990s is on performance and creating high performing organizations, the focus in the 2000s has been on producing socially sustainable vertical relations. The current economic crisis has, at least to some extent, brought back concerns with organizational effectiveness and the organizations' ability to produce high performing organizations in an efficient manner. These particular concerns also seem to permeate the report made by the Productivity Commission (2013) on effectiveness in the Danish Public Sector. In chapter 5 of their report, they point to that

in order to secure effective vertical relations which also supports the effectiveness of the organizations a higher emphasis on the relation between individual performance and pay must be established. Like the theories I will discuss in chapter II, the Commission also advocates the active use of pay to support quantitative optimization with a focus on effort and qualitative optimization with focus on developing the individual employee's competences and abilities in a manner which supports the organization's development. Hence, the main argument seems to be that in order to produce high efficient organizations, one has to identify the employees as the primary source – as the source of quantitative performance as well as the source of the qualitative foundation in terms of the organization's competences. Within this perspective, the primary source of optimization is the vertical relations as also advocated in contemporary theories on compensation. However, whereas it seems undoubtable true that the primary source of effectiveness is human resources, it might be questioned if this focus does not neglect the relational perspective which seems inherent to understanding organizations as social systems – and not solely in a vertical manner, but in a horizontal manner, in the sense that whereas results may be brought about by the performance of individuals, this performance unfolds itself within a social context and, as such, social relations may not be viewed as the primary source of production, but must be viewed as producing the *social fabric* of the organizations, the fabric which enables it to function, the fabric which prevents social interactions between employees and across hierarchical boundaries from turning into endless counterproductive disputes. In that sense, the continuity and functionality of horizontal and vertical relations enables the organization to function by producing the social framework of social interaction. From this perspective, then, the question becomes a matter of understanding how these social relations are maintained and secured and the notion of social rationality, introduced above, may provide some insight into this, in the sense that this principle derived from rational philosophy explicates the fact that rational governance of one's life necessitates not just a focus on effectiveness and limits associated with this, but also limits associated with the social world in terms of, for example, the fact that actions has social consequences which the acting individual needs to take into consideration. From a less abstract perspective, it might also just be argued, as one of the interviewees did in my empirical study, no one likes being disliked by others. This notion of preferences regarding how other individuals perceives one

may also be associated with what is often, in modern cognitive psychology, called the need to belong – the need to be a part of something which transcends the immediate present and material life. As I will argue in chapter VIII, this need to belong is necessary to secure if the organization wants to produce highly motivated employees – employees motivated to develop themselves, employees who are motivated by a desire to realize their potential and not just seeks monetary means to satisfy their basic needs. Hence, securing these social relations becomes a matter of pragmatic necessity, however, not simply in order to secure satisfaction of this basic psychological need, but also because it is from these relations that claims and obligations emerge.

From this perspective it also seems to follow that actions are not caused, solely at least, from the outside, just as it seems to follow that while the motivation behind an action may be purely egoistic, the action must be of such a kind that it does not damage these social relations. What is argued here, then, is not that social agents are not or should not be egoistic or that compensation systems which motivating egoistic behavior will not function; rather what is argued is simply that egoistic behavior coexists with non-egoistic behavior and that a compensation system which destroy these foundational social relations might end up undermining its own functionality, in the sense that individuality and individual performance only unfolds itself and flourish within stable social environments and withers within environments characterized by war-like competition similar to the dystopian state of war described by philosopher Thomas Hobbes (1651/1985). One needs, in other words, to understand how actions are produced within a social world which is horizontally as well as vertically structured and that these actions are produced to unfold themselves within this context. Moreover, these actions are assumed to be produced by rational decision-makers. The effective level of performance is not, solely, that which maximizes the outcome and minimizes the costs of production, but that which do so while still maintaining the productive foundation which cannot be reduced to individuals but must, at least to some extent, encompass social relations. The main question, in accordance with this, becomes not *just* how the organization most effectively motivate its employees to supply high levels of performance or develop themselves in accordance with the organization's needs, but also how it does so in an internal sustainable manner – a manner which maintains the social relational foundation.

Since the pivotal point of this work is the necessity of social relations, it seems only fair that I acknowledge those who have made it possible for me to write this work. First of all, I wish to thank the companies and the respondents who participated in my empirical study. But most of all, I wish to thank my supervisor Steen Scheuer for his support, motivation, patience, and not at least for always forcing me to be more precise, more explicit, or simply – better. I also wish to thank Anders Bordum for his always critical and sharp comments, support, friendship, and for introducing me to the academic world; Peder Madsen without whose help I would not have become a PhD-student; and my colleagues Kristian Rune Hansen and Eva Riis who I have had the pleasure of sharing an office with and who have listened, patiently, to my worries. Finally, I wish to thank my family for their love and support, just as I wish to apologize for my absent-mindedness and, at times, intolerable behavior during the last three years.

## Problem Statement

Reflecting on these matters, described in the foreword, related with understanding fairness as related not simply with consequences, but also with actions emerging from horizontal and vertical relations leads to a causal perspective on compensation, just as it leads to a focus on how these causal processes unfold within a social context and are influenced by the normative foundations inherent to the basic social relations. The primary aim of this thesis, then, becomes to understand how compensation strategies framed as causal processes which unfold themselves within vertical and horizontal relations are influenced by the normative foundations inherent to these relations. That is, fairness is not just a matter of evaluating whether or not some outcome is in accordance with the normative expectations inherent to vertical and horizontal relations, it is also a matter of evaluating whether the actions which brought about these consequences are in accordance with the relational normative foundation. Focusing on the relational structuring of fairness as associated with the normative constitution of claim rights leads to a focus on how these might interact with the values on individuality and individualization inherent to contemporary compensation theories, just as it leads to a focus on how this interaction might influence the horizontal relations in the organization, in the sense that the outcome may be vertically defensible, but the actions needed to bring it about may not be horizontally defensible – a conception which follows from the causal structuring of the outcome situation. The primary research question can now be formulated:

How is the fair claim constituted in, respectively, a compensation system based on individualization and individuality, a social relational perspective; and in which ways might potential differences influence horizontal relations?
--

The primary research question will be analyzed based on following three research questions:

- 1) Analyze and discuss the constitution of the fair claim in contemporary theories on compensation systems based on the two values individuality and individualization. That is, based on an exploration of the particular conception of fairness

which follows from these two values, discuss and analyze what constitutes a fair claim – a claim which is defensible and acceptable?

- 2) Analyze and discuss what constitutes a fair claim from a relational perspective. That is, what characterizes a fair claim horizontally as well as vertically and how is it constituted?
- 3) Based on research questions 1 and 2, analyze and discuss how potential differences between how the fair claim is constituted in contemporary compensation systems and how it is relationally constituted might influence the horizontal relations in the organization.

# Chapter I

## Reflections on the Philosophical and Theoretical Foundation

### 1.1 Introduction

The aim of this chapter is to discuss the philosophical foundation of this thesis – that is, to discuss the epistemological, ontological, and theoretical assumptions on which this thesis is based. The implications of this in relation to methodology will be discussed in chapter IX.

### 1.2 Knowledge and Theoretical Construction

The aim of this section is to discuss the epistemological and ontological foundation of this thesis. Ontology, here, refers to assumptions on the state of what is claimed to exist, whereas epistemology, here, refers to assumptions on how knowledge is gained about these states. Ontology is here not applied in its conventional sense as assumptions about that which exists, rather as assumptions on the nature of that which exists, and as such it is closely connected with epistemology – the way knowledge is gained about these states, but also the layers of knowledge which goes into the epistemic construction of that which is claimed to exist (Bhaskar, 1986/2009: 6; Danermark, Ekström, Jakobsen, and Karlsson, 2002: 18).

Inquiring into the state of being also implies a process perspective on the ontological dimension – that is, the object of the ontological inquiry cannot be comprehended in its essence, insofar as its essence cannot be captured in itself, rather the object must be perceived as consisting of different layers and relations (Sayer, 1992: 62). Relations describes the constitution of the object – the composite nature of the composition – or the relations between the differentiated elements, while layers refer to the internal composition of the object – that is, the object of an ontological inquiry can be horizontally differentiated into different elements, just as it can vertically stratified into different layers (Bhaskar, 1978/2008: 15, 29, 87, 169). The vertical perspective can be associated with the choice of analytical level – level of aggregation – while the horizontal perspective can be associated with the choice of elements allowed into the analysis. In that

sense, the state of the object changes as one shift between different states of the object – different levels of abstraction, horizontal as well as vertical (Sayer, 1992: 86; Danermark, Ekström, Jakobsen, and Karlsson, 2002: 42). The simultaneous occurrence of these two dimensions may be characterized as the event which unfolds itself across the spatiotemporal frame implying that the apprehension of this phenomenon cannot be captured in its perfection at one time, insofar as it is always changing – it is in a state of flux. This domain of reality may be called the actual domain and can be distinguished from the empirical domain – the domain at which sense and experience exists. The third domain is the domain of the real the deepest dimension of reality at which the generative mechanisms exist (Bhaskar, 1978/2008: 56; Danermark, Ekström, Jakobsen, and Karlsson, 2002: 20-21).<sup>1</sup>

A science reduced to the domain of empirical reality, reduces reality to that which is immediately present to the perceiving agent – that is, reality is collapsed into one domain, and as such it is assumed that reality can be apprehended in its perfection at one moment – the temporally and spatially closed observation of reality as it is in itself – the apprehension of actuality – that which is actually or immediately (here and now) present and can be perceived through the perceiving agent's sensory experience (Bhaskar, 1978/2008: 58; 1986/2009: 23, 28, 31). In that sense, the epistemological point of departure, is the experience of the perceiving agent which entails that epistemology as well as ontology is anthropocentric in origin. However, insofar as the object cannot be fixed in the immediate present, but unfolds itself in the simultaneity of the horizontal and vertical dimensions, moving from the domain of empirical reality to the domain of the actual also entails a move from a static perception of objects to the state of occurrence (Bhaskar, 1978/2008: 78; 1986/2009: 27). Moving from the state of occurrence to the domain of reality implies moving from what occurs to the generative force behind that which occurs (Bhaskar, 1978/2008: 92). That is, transcendental realism implies a move from the reality as it unfolds itself to the level of the generative mechanisms be-

---

<sup>1</sup> Steve Fleetwood (2005: 199-202) presented a perhaps less abstract notion of ontology based on a pragmatic criteria of influence – although something may not exist as a material entity it may still have an impact, that is, the real is that which has practical influence. Furthermore, he distinguishes between four modes of reality: *Material real entities* referring to objective objects whose existence is independent of the observer; *Ideally real* refers to discourses, language, ideas, representations, beliefs, opinions, etc.; *Artefactually real* entities referring to objects whose meaning is entirely conceptually dependent such as, for example, computers; and *Socially real* entities such as organizational structures.

hind that which occurs (Bhaskar, 1978/2008: 49; 1986/2009: 31; Sayer, 1992: 104). From a *Popperian* perspective, the move from the level of the empirical and the actual, to the level of the real is associated with the identification of *universal* dispositions inherent to the object – that is, the behavior of the object under certain conditions, if, for example (to borrow an example from Popper) “breakable” is a disposition inherent to the object, the characteristic of the object’s behavior under certain conditions – its ability to break – is a reference to or an assertion about, the character of that object, the *expected* behavior of that object under certain conditions which are due to its inherent qualities (Popper, 1963/2002: 158, 159). I return to the distinction between mechanisms and universal laws below. The empirical and actual reality can, from a Popperian perspective, be viewed as the dimensions of the object which have extension in time/space, and the non-extensional dimension related with pure appearance – for example, color (Popper, 1963/2002: 155). Furthermore, Popper’s (1963/2002: 157) main argument is that reality cannot be identified drawing on a distinction between the real and the unreal if this is based on a distinction between that which has extension and that which does not, and neither can it be based on a distinction between that which is immediately present to one’s senses and that which is not, insofar as reality has more than one layer – implying that the distinction is based on another distinction, induction/deduction – therefore, the layers which only emerge through the use of deduction are not *less* real. Bhaskar’s (1978/2008) critical realism can, then, be viewed as a reformulation of this Popperian perspective, in the sense that he tries to push past the *Kantian* rationalism inherent to Popper’s perspective – hence, instead of the transcendental idealist distinction between “the world as it is” and “the world as it appears”<sup>2</sup>, critical realism poses this as a matter of the social construction of knowledge – that is, the cognitive is trans-

---

<sup>2</sup> According to transcendental idealism, one may distinguish between the world of appearances and the world of objects in themselves. The latter world is characterized by the dimension of the objects which exists independent of the perceiving agent, while the former world consists of the dimension of objects whose properties are contingent upon the perceiving agent. Therefore, one may distinguish between a world consisting of external objects and a world consisting solely of mental representations of these objects – that is, how these present themselves in the mind of the perceiving agent as a kind of phenomenological apperception. Kant’s perspective is described in the following quotation: “...all our intuition is nothing but the representation of appearance; that the things which we intuit are not in themselves what we intuit them as being, nor are their relations so constituted in themselves as they appear to us, and that, if we remove our subject or even only the subjective constitution of the senses in general, then the entire constitution and all relations of objects in space and time, nay space and time themselves, would vanish. They cannot, as appearances, exist in themselves, but only in us.” (Kant, 1781/1787/2007: A42, 43/B59, 60; 75)

formed from a matter of psychological/mental faculties to a matter of social processes which influences human apperception.

### **1.2.1 Causal Mechanisms and the Dimension of the Real**

The level of reality associated with these generative mechanisms is the most general level of explanation in the sense that it implies that the researcher takes one step back to discover the transfactual *laws* which exists independent of the researcher, just as they exist independent of the spatiotemporal conditions (Bhaskar, 1978/2008: 52). It entails, in other words, a move from actuality to potentiality – a move from that which is currently unfolding itself (actualizing or being actualized) to the conditions which determines the possibility is activated in the object, although the transcendental realist may not have as it aim to discover the conditions of possibility, but rather focus on the *normic* nature inherent to the object – a tendency which might be activated or a potentiality (a power) (Bhaskar, 1978/2008: 96, 98, 225; Danermark, Ekström, Jakobsen, and Karlsson, 2002: 55). In that sense, a normic statement may be perceived as a statement about a frequently occurring and easily recognizable causal pattern (Sayer, 1992: 107; Bhaskar, 1978/2008: 221). Although the pattern is frequently occurring and easily recognizable, it is impossible to identify when it is activated and to predict its results (Bhaskar, 1978/2008: 99; Sayer, 1992: 108; Danermark, Ekström, Jakobsen, and Karlsson, 2002: 52). As such, it may be understood as a possible explanation of what occurs with reference to the potential of the object – it represents a possible and probable explanation of the event's occurrence. In that sense, the aim of such an inquiry is to tap into the powers or liabilities of the object of interest and its constitution – that is, one must distinguish between possibilities, limitations, and weaknesses of the object in the sense that the object consists of properties specifying the potential of the object – its possibilities and tendencies – the limitations of the object – the space of possibilities is not infinite – and weaknesses which might influence the potentiality of the object (Bhaskar, 1978/2008: 221; Sayer, 1992: 103, 105; Danermark, Ekström, Jakobsen, and Karlsson, 2002: 57). For example, in employee-colleague relations, the employee may have a specific set of rights inherent to the relation between the parties, this right may however not be exchanged into other rights – extending the obligation of the other – just as her actions may reduce the causal efficacy of the rights – her rights to claim – insofar

as transgressions may reduce the likelihood of her colleague accepting the obligations inherent to the particular claim (see also the discussion of claims and rights in chapter III).

These causal mechanisms which are currently invoked as possible explanations of the occurring phenomenon may be described as the intransitive dimension of reality – that which is the result of something preceding it – its antecedent (Bhaskar, 1978/2008: 66). The transitive dimension of reality explains that which is occurring, and as such it is the theories invoked in order to explain this – that is, the activated frames of interpretation. The move from the intransitive to the transitive entails, then, a move from ontology to epistemology. The transitive dimension, in other words, is the ready-at-hand explanations of that which occurs, thus it cannot be reduced to scientific or common-sense knowledge (Sayer, 1992: 16); rather it is a mixture of these insofar as it describes the continuous pursuit of meaning (Danermark, Ekström, Jakobsen, and Karlsson, 2002: 27, 35). As such, the transitive dimension provide the link between science and reality in the sense that these theories expresses accepted and ready-at-hand explanations, albeit in a limited sense insofar as these theories are neither complete (infallible) nor independent of practice – the rules (guiding the process of sense making) and reasons (problem statement etc.) (Bhaskar, 1978/2008: 60; 1986/2009: 51, 68; Danermark, Ekström, Jakobsen, and Karlsson, 2002: 23). These already existing theories, then, expresses the epistemological point of departure, just as they express what is currently known about the object of interest (Bhaskar, 1986/2009: 63, 65). Following this, it might be argued that the object which appear in front of the subject is already framed in a particular way by the existing system of knowledge on which the researcher draws in her process of sense making – that is, what is known is pre-structured by existing theories and the concepts invoked to explain the phenomena of interest and the causal relations described (Bhaskar, 1986/2009: 25). Hence, it follows from this that apprehension is mediated by existing theories, although it does not follow that reality is determined by these existing theories (Fleetwood, 2005: 199).

Theories, then, are cognitive maps which structures the process of sense-making, although this should not be interpreted in a rigid sense, insofar as research is not reduced to a passive reproduction of these existing frames of interpretation; rather these

existing theories and practices, within which science unfold itself, should be viewed as constituting the conditions of possibility in the sense that these along with the abilities of the researcher determines the properties of the results (Sayer, 1992: 20). That is, the process of actualizing the potentials in the horizon of possibilities is contingent upon the researcher's abilities, creativity, etc., just as it is limited by limitations inherent to practice – the external limits of practice associated with, for example, intersubjective evaluation inherent to the practice of science (Bhaskar, 1986/2009: 52, 65; Danermark, Ekström, Jakobsen, and Karlsson, 2002: 29, 37; Sayer, 1992: 20). Here it seems, at least to some extent, that critical realism is approaching the metatheory of scientific explanation similar to those proposed by, among others, Lakatos' (1970: 191) idea about the heuristics of research programs and Kuhn's (1962/1996) scientific sociology, and his concept of paradigms as research programs, although the meta-philosophical approach applied by critical realists aims at identifying the conditions of possibility of science from an ontological dimension in terms of a knowledge problem.

In addition to this, it might be argued that the context of understanding is not the context in which the object is situated or appears; rather the context is a social context in which the subject finds herself – that is, the object is not understood as being the product of a particular social context; rather the context is the epistemological frames inherent to the system of knowledge on which the observer draws in her process of sense-making (Sayer, 1992: 21). Therefore, the meaning of the object is not dependent on the context in which it is situated; rather the meaning of the object is dependent upon the epistemological (explanatory) context on which the observer draws. That is, the object does not become a particular object by reference to the particular horizon (context) on which it exists; rather, in what seems to be a Nietzsche (1873/1976: 45) inspired critical perspective, reality is impregnated with meaning through the meaning which is projected on to it by the perceiving agent – this meaning, however, is not subjective in nature insofar as it is based already existing frames of interpretation, just as it is formulated in language, and as such determined by the conventions and rules on which the linguistic practice is based. These already-existing theories along with language determines the conditions of possibility of the frame of interpretation in a manner which also limits the *artistically freedom* of interpretation, inasmuch as the interpreter who deviates from these existing theories must intersubjectively justify her perception, just as deviations

would demand that the perceiving agent was able to put on hold these existing frames – and by doing this she would find herself confronted with the almost impossible task of defining reality on her own terms (Sayer, 1992: 29; Nietzsche, 1873/1976: 47).

In continuation of this, the epistemological dimension cannot be reduced to sense data insofar as the aim of transcendental realism is to push past reality as it unfolds itself in the immediate present moment, just as its aim is also to push past mere descriptions of what occurs and patterns of interpretations; rather the aim is to explain the general tendencies, the generative mechanisms, that which explains the relation between the context and the event which unfolds itself in practice (Danermark, Ekström, Jakobsen, and Karlsson, 2002: 36, 37). The inclusion of practice implies that epistemology cannot be separated from ontology – in the sense that practice is derived from understanding, different ways of understanding may, therefore, come to be expressed in different manners of performance. In other words, practice as a reflection of activated cognitive schemata also reflects the psychological as well as social generative mechanisms of reality in its *real* sense. In other words, applying this perspective entails a particular understanding of judgment as a social practice – a practice which is not just influenced by the activated patterns of interpretation, insofar as common patterns of judgment must reflect more than just shared frames of interpretation – they come to reflect the conditions of the practice of passing the judgment (Bhaskar, 1986/2009: 17). The aim of transcendental realism, then, is to inquire into these conditions of production in the sense that these constitute the intransitive dimension of reality – what is presupposed about the world in the unfolding practice (Bhaskar, 1986/2009: 60; Danermark, Ekström, Jakobsen, and Karlsson, 2002: 44). These generative mechanisms cannot be observed; only their effects can (Danermark, Ekström, Jakobsen, and Karlsson, 2002: 43). The question may then be – how are these mechanisms to be determined?

First of all, by postulating the existence of some mechanisms is also to postulate the existence of some cause which precedes the result observed – that is, the result observed belongs to the level of empirical reality, whereas the event which this is part of belongs to the level of actual reality, and the cause belongs to real domain of reality. A transfactual approach to science, then, aims at producing the necessary connections between these three different levels of reality (Danermark, Ekström, Jakobsen, and Karls-

son, 2002: 57). Following the approach described here, the event which is currently unfolding itself must be viewed as part of a longer causal chain unfolding itself. The aim, however, is not to identify the starting point of the causal chain which must be assumed; rather the aim is *simply* to identify the mechanism which allows one to explain the transformation of a cause into the observed effect (Bhaskar, 1989/1998:52; 1986/2009: 131; Hedström, 2005: 25). That is to say, the aim is not to identify the connection between two discrete events (Sayer, 1992: 104; Danermark, Ekström, Jakobsen, and Karlsson, 2002: 54); rather the aim is to explain the process which enables the transformation. From this it also seems to follow that the theory chosen to explain the observed will explicate different points in the causal chain, in the sense that it explains different levels (or strata) and different causal relations by identifying different causes all belonging to the same causal chain (Bhaskar, 1989/1998: 43; Sperber, 2011: 71; Danermark, Ekström, Jakobsen, and Karlsson, 2002: 49, 59; Hedström, 2005: 32).

To explain a certain social phenomenon, one may draw on different explanations, for example, from a sociological point of view, one may explain judgment as a matter of culturally embedded frames of interpretation on which the individual may draw; whereas from a cognitive psychological (psychosocial) point of view one may explain judgment as a matter of the relation between input (or stimuli), perception, and outcome (mental representation) – that is, for example, judgment may be explained as a matter of the frame on which the individual draws when she tries to make sense of the input in order to produce a *right* and *truthful* mental representation of the input; just as one may also explain judgment as a matter of biological or neurological factors influenced either by, for example, evolutionary biological factors or neurological factors related with the stimuli produced by the brain. None of these explanations can be viewed as dominant, insofar as they just describe different ways of explaining the same phenomenon, and as such they explicate different levels of the causal chain, just as they draw attention to the fact that reality is simultaneously neurological, biological, and social (see, for example, Elster, 2007 for a *catalogue* of different mechanisms which might explain individual as well as social action; Bhaskar, 1978/2008: 222; 1986/2009: 125; Danermark, Ekström, Jakobsen, and Karlsson, 2002: 45). This does not, however, entail that one should always include all possible strata to explain a particular phenomenon, insofar as it may

not make sense in relation to, for example, the aim of the research program or problem statement (Sayer, 1992: 120; Danermark, Ekström, Jakobsen, and Karlsson, 2002: 63).

The different perspectives, then, can be viewed as different causal stories of the same phenomenon. On the surface (or the denotative level) a particular action may have several meanings – in contrast to Clifford Geertz (1973) (or Gilbert Ryle's) interpretation of thick descriptions focusing on the relation between the context and the action, one may, from a somewhat different perspective, focus on meaning as a product of different causal chains – that is, the difference between one explanation of a particular gesture and a different explanation of the same gesture is the causal history invoked to explain it (Sperber, 2011: 72). A thick description, on this account, then, entails that one tries to *flesh out* or unfold the causal history of the phenomenon of interest in order to explain the transformation of the cause into the effect observed (Hedström, 2005: 31). Social phenomena may, based on this, be explained as the intersection of different individual causal chains, that is, actions (or non-actions) of one social agent becomes input into the other social agent's causal chain, therefore social life may be described (in a simplified form) as a series of intertwined causal chains (Bhaskar, 1986/2009: 129; Sperber, 2011: 70). What matters is, in other words, the causal story attributed to the event in the sense that the event comes to be viewed as a result of the causal chain attributed to it. This attribution of causal history is, at least to some extent, a cognitive process in the sense that it entails cognition (Bhaskar, 1986/2009: 119, 134; Sperber, 2011: 72). A cognitive process which is not decoupled from its biological/neurological or its social foundation – that is, the mental representation is influenced by internal (biological/neurological) as well as external (social) factors (Sayer, 1992: 119). As such, the aim is to produce a causal story which enables one to explain the occurring event – the mechanisms which allow one to explain the results as a matter of these triggered mechanisms. The mechanism which is postulated to explain the particular social phenomenon should, then, produce a *plausible* story which is consistent with the information (input) one has, in the sense that it must be consistent with the causes and the effect, hence the space of possible mechanisms from which one can choose is limited by the information transferred by the occurring event (Danermark, Ekström, Jakobsen, and Karlsson, 2002: 45).

A mechanism is not a description of what actually occurred nor is it an attempt to explain the complex process of choice which preceded the actual choice; rather it can be viewed as the *postulation* of a general explanation which allows one to explain the event – that is, if one observes that several respondents chooses, for example, a particular distribution in a dictator game, the aim is not to explain what made the single individual chose that particular distribution, insofar as the answer would be too complex to capture, because it might draw on several factors which cannot be separated analytically nor can it be assumed that the respondent is actually conscious of her reasons (Danermark, Ekström, Jakobsen, and Karlsson, 2002: 62). Explaining the phenomenon by indicating the existence of a particular mechanism or several mechanisms, then, implies that one shift from the actual event to a more general level and describes the nature of mechanism – when it is *normally* triggered and with what results (Bhaskar, 1978/2008: 222; Sayer, 1992: 104, 111; Danermark, Ekström, Jakobsen, and Karlsson, 2002: 56; Hedström, 2005: 11). The mechanism cannot, as such, be viewed as an explanation of why a particular social agent acted the way she did – all the mechanism claims to explain is why such actions, at a general level, are performed and with what result (Danermark, Ekström, Jakobsen, and Karlsson, 2002: 50).

A mechanism – a general tendency – then, is not a general/universal law, although it shares with it, the characteristic of being general in the sense that it is not limited by subjective characteristics (for example psychological state) nor spatiotemporal frames (see, for example, Popper 1957/2002: 106; 1935/2002: 40). From this it also follows that a mechanism is not a *trend*, insofar as a trend refers to statements of something which occurred at some time at some particular place (Popper, 1957/2002: 58, 106). According to Popper (1957/2002: 106; 1935/2002: 47) a trend is purely existential, because it is based on the assertion that such-and-such exist – was present or occurred at some time and place – and as such it is purely metaphysical in nature. Rather, a mechanism may refer to the qualitative explanation of a trend, manifesting itself as quantitative patterns – a mechanism, therefore, refers to another level of analysis than trends do (Boudon, 1976: 1187; Hedström, 2005: 30). Moreover, to postulate the existence of a mechanism is not to postulate the existence of a causal explanation – at least not as based on a deductive logic (as analytic sociologists would) – taking as point departure some universal law and combining it with a specific characteristic of a certain particular

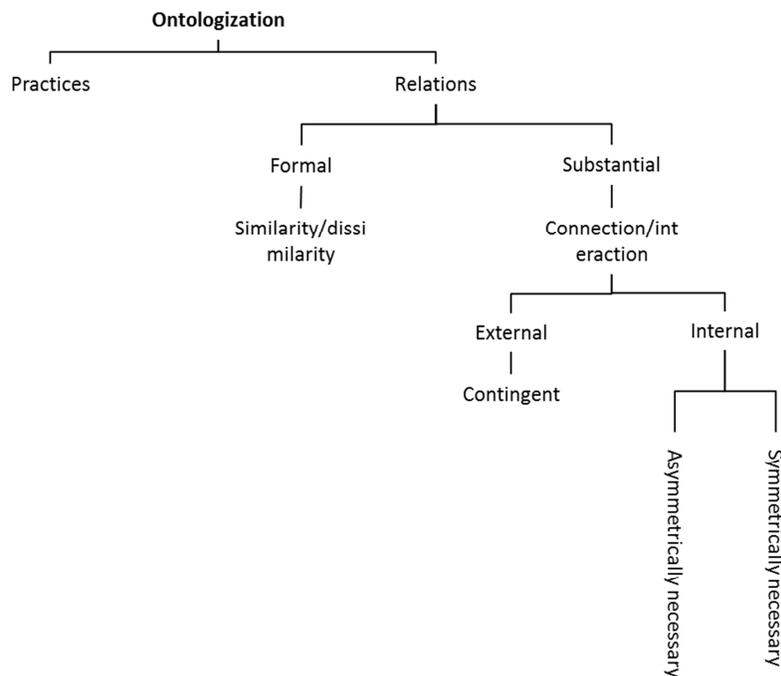
event (Popper, 1935/2005: 38; 1957/2002: 113; Bhaskar, 1989/1998: 45). The primary problem associated with such a definition of causality is that it presupposes the existence of testable conditions which again presupposes the existence of a closed system – that is, it presupposes the possibility of isolating causes and effects which can be formulated as a hypothesis and tested. From a Popperian perspective, it might be argued that the problem with explanations of this kind, is that it is based on certain perspectives (interpretations) which cannot be tested, whereas this, from a critical realist perspective, poses no such problem, insofar as the progress of science is not based upon producing temporarily unfalsified theories, but rather, *simply*, better explanations and explanations of a more general kind (Popper, 1963/2002: 1957/2002: 139; Bhaskar, 1989/1998: 46).

Rationality, for example, may be indicated as a particular mechanism which enables the researcher to explain a particular course of events (see, for example, Coleman, 1990 or Hedström, 2005). Emotions understood as cognitive orientations, norms, empathy based on, for example, identification (cognitive) or mirror neurons (neurology), etc. are other examples of mechanisms which might be invoked to explain various social phenomena (see, for example, Elster, 2007: chapters 8 and 17). Additionally, explaining by mechanisms also allows one to explain social action without addressing issues such as, for example, intentional/non-intentional results or consciously/non-consciously produced actions, although these social phenomena are important, it is problematic to identify them without assuming something which might not be supported by the evidence one has – for example, claiming that a result deviated from the intended results implies that one assumes the existence of a different and intended aim (Bhaskar, 1986/2009: 127-128).

From what has been argued so far, it might be argued that a particular social phenomenon can be explained drawing on different theories each addressing different parts of the causal history from different perspectives and levels. However, because a social phenomenon can be viewed as the simultaneous occurrence of several sub-level phenomena which combined generates the actual event, it is crucial to specify the phenomenon of interest – that is, the particular dimension of interest (Danermark, Ekström, Jakobsen, and Karlsson, 2002: 46). The specification of the phenomenon of interest allows one to *ontologize* on the phenomenon – that is, transform it into an object de-

scribed by a particular *structure* consisting of a particular set of relations and practices (Sayer, 1992: 88; Danermark, Ekström, Jakobsen, and Karlsson, 2002: 45). According to Sayer (1992: 88) it is possible to distinguish between relations which are *substantial* of connections and interaction; and *formal* of similarity or dissimilarity. Furthermore, according to Bhaskar (1989/1998: 42) it is also possible to distinguish between internal and necessary; and external and contingent relations (see also Sayer, 1992: 89). A is internally related to B if and only if A's existence presupposes the existence of B – if this also applies for B, the relation is also symmetrical (Bhaskar, 1989/1998: 42). A is contingent on C if the relation is neither impossible nor necessary – that is, C may influence A, but it is not necessary for the existence of A (Sayer, 1992: 89). In addition, from what was argued above on the nature of mechanisms as possible and plausible explanations, it seems to follow that the relation between the mechanism and its effect is external and contingent insofar as the mechanism may not be activated; whereas, on the other hand, the relation between the object and its powers, liabilities, and weaknesses are internal and necessary (Bhaskar, 1978/2008: 221; Danermark, Ekström, Jakobsen, and Karlsson, 2002: 55-56). This might be illustrated as follows:

Figure 1: The process of ontologization (inspired by Danermark, Ekström, Jakobsen, and Karlsson, 2002: 46)



Based on this, a tentative description of the social phenomenon of interest here can now be made. From the problem statement, it follows that one can distinguish between a horizontal and a vertical dimension of internal necessary to social relations – the vertical dimension referring to the relations: organization-employee and employee-manager; and the horizontal dimension referring to the relation: employee-colleagues. In particular, the aim is to explain how the normative expectations inherent, primarily, to the horizontal relational dimension might influence judgments on compensational fairness, and following upon this, how these judgments may influence the social relation employee-colleague when transformed into cognitive orientations expressed in interaction. Furthermore, to explicate the causal nature inherent to compensation, one may distinguish between direct and indirect causes of relational changes stemming from compensational changes – that is, the direct cause of a compensational change may, for example, be the decision of the management to reward a particular employee, whereas the indirect cause is the actions of the receiving individual. The direct cause may also be viewed as exogenous to the relation employee-colleague, while the indirect cause is endogenous to the relation insofar as one of the agents acts in a way which sets in motion the causal chain (see chapter II). The basic argument proposed here is, then, that judgment on fairness as well as the relational influence is influenced by both matters – the normative expectations inherent to the relations and the causal nature of the compensational process. In that sense, the aim is to analyze and discuss how normative expectations may influence judgments on fairness, and how these judgments may interact with the relational effects of compensational differences. Furthermore, it must be assumed that different psychological needs, for example, the need to belong may influence the potential of the agent insofar as she might not act in ways which she perceives as undermining of this psychological need (see chapter VIII). In that sense, what will be argued is that fairness can be viewed as the foundational threshold of acceptance, in the sense that what is fair must also be acceptable – the question then becomes a matter of identifying the constituting, social as well as the cognitive, conditions of fairness as acceptance.<sup>3</sup> Fairness as acceptance, however, cannot be viewed as reduced to the vertical relations, insofar as fair-

---

<sup>3</sup> A similar approach to judgment on moral questions was proposed by Kant (1781/1787/2007: A44, 43/B61, 62; 76) when he argued that in regard to morality it makes no sense to consider the constitution of nature; rather, what matters is the moral constitution. Both phenomena, however, are contingent upon the agent's cognitive representation, although Kant would accept the influence of the context upon judgment; see also the discussion of Kant in chapter III.

ness as acceptance also implies acceptance of compensational differences which seems to point to another crucial question: why is it that compensational differences are accepted? This question is related with the horizontal dimension, and as such it suggests that judgments on fairness are related with normative expectations inherent to this relation. The primary aim, in other words, is to explain how fairness judgments *might be* formed – that is, which factors might influence the judgment and with what results.

Based upon the discussion above, one might, inspired by Bhaskar (1978/2008), ask in which world is it possible to perform such an inquiry – that is, on which assumptions, epistemic and ontological, is such an inquiry centered? Firstly, it is assumed that fairness judgments are influenced by normative expectations inherent to vertical as well as horizontal relations; secondly, it is assumed that social agents are relationally bound to one another, and that these relations, as well as the normative foundation on which they are centered, influences judgments on fairness, insofar as these are constitutive of rights and obligations; thirdly, it is assumed that social agents have social needs which are psychologically nested as well as emotions manifested as cognitive orientations towards oneself and others; fourthly, it is assumed that compensation systems exist as a set of values and expectations which guide the employees decisions on various professional areas, just as it influences the normative foundations of the two relations; and fifthly, it is assumed that different social phenomena such as, for example, a particular reward practice or compensatory differences are ascribed certain social meanings due to the relation between these and the normative expectations inherent to the two relations, just as they might come to influence the future state of these relations – that is, the cognitive orientations which accompanies these.

The normative expectations inherent to the two relations can be viewed as partly local and partly global, in the sense that the *local* structuring of the relations may deviate from the *global* frames inherent to, for example, an Anglo-Saxon or a Scandinavian conceptualization of the horizontal and vertical relations. Nevertheless, the normative expectations, although individually represented, are socially constituted in the sense that these structures are not the creation of the individual agent; rather they are shared frames of expectations, rules, etc. in the sense that they must be viewed as pre-existing, and as such, they generate the limitations of the space of possible actions present to the

agent (Bhaskar, 1989/1998: 36, 38; 1986/2009: 122, 127). Here it should be remembered that effects are solely externally linked to the mechanism which implies that these structures can primarily be viewed as suggestions or possibilities, hence they can never be viewed as the direct or single cause, insofar as this would entail a deterministic approach; rather one should perceive human action as being a result of an intentional and conscious process accompanied by the *unconscious* reproduction of the conditions (Bhaskar, 1986/2009: 135) – the limits which constitute the space of possible actions – like the linguistic practice, it is limited by pre-existing frames, which simultaneously enables the practice, but they are not determined by these (Bhaskar, 1989/1998: 35, 36; 1986/2009: 123). Applying this conceptualization of normative expectations also implies that these are assumed to have a pragmatic impact on the actions of the individual agent entailing that they are assumed to exist independently of the social agents – therefore, they may be viewed as ontologically objective (Bhaskar, 1986/2009: 124). Even though social structures can, in an analytical abstract perspective, be viewed as ontologically objective, they cannot be viewed as epistemic objective, insofar as the knowledge of these is, *undeniably*, individually manifested – hence, the epistemic nature of social structures is subjective (Bhaskar, 1986/2009: 131). That is, an inquiry into the nature of these normative expectations must take as point of departure, the manifestation of these in, for example, statements or the knowledge of individuals – lay knowledge. These normative structures, although they are assumed to exist independently of the agents, cannot be accessed independently of the agents whose practices are influenced by them. These structures, nevertheless, may be accompanied by a claim on generality manifested in claims on, for example, *acceptance* or *rightness* accompanying actions in the form of tacit/implicit statements like, for example, “if another individual (generalized other) had found herself in a similar situation, she would also have acted in a similar way as I did” (see the discussion of Raymond Boudon in chapter III). Of course, the action need not be accompanied by this claim on generality, insofar as the reasons behind the reproduction of normative structures may also be due to, for example, the grip these have on the mind through the relation between emotions and normative structures (see the discussion of Jon Elster in chapter VIII).

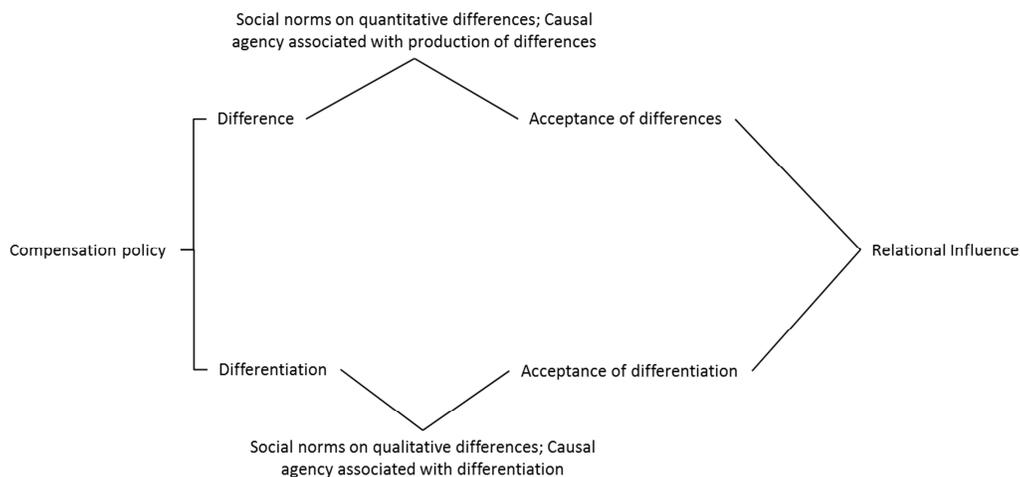
The paradoxical nature of these social phenomena can also be viewed as a result of the particular nature of social phenomena, in the sense that these are socially consti-

tuted epistemologically as well as ontologically (Bhaskar, 1986/2009: 133). In that sense, a transcendental inquiry into the *essence* of a social phenomenon is, at least, problematic due to the particular nature of this unless one, of course, adopts an *externalizing* approach which may, at least to some extent, reify and mystify the existence of these social phenomena (Bhaskar, 1989/1998: 33). The essence of these social phenomena is, in other words, the knowledge/theories held by these individuals whose actions reproduce or transform them, hence they gain pragmatic impact from the fact that they are taken into consideration, just as the claim on generality which accompanies them (Bhaskar, 1989/1998: 39; 1986/2009: 124, 165; see also chapters III and VIII). To some extent, then, critical realism takes on a middle position on the question of structures, in the sense that it argues that these structures are constitutive of and constituted by practice, however, not in manner which imply structural determinism; rather these structures exist independently in a cognitive manner as *self-imposed* limitations, and as such, the perspective can be distinguished from the dialectical perspective associated with some forms of social constructivism (see, for example, Berger and Luckman, 1966/2002:79). That does not, of course, imply consciousness of these conditions, but it does entail that reason is the cause of actions – that is, actions are manifestations of reasons in the sense that the agent does what she wants (Bhaskar, 1989/1998: 95; 1986/2009: 167). A reason is a belief – a belief on powers (space of possible actions social and physically constituted), what she has done, what she has been able to do, what she has desired, the desired states she has been able to bring about, etc. (Bhaskar, 1989/1998: 95; 1986/2009: 135, 160). As it was also argued above, action may be, in some sense, initiated as a response to some other agent's actions – however, a response requires consciousness and intentionality in form of choice of appropriate response. An action can, in that sense, be caused by some external factor, albeit this cause is not the primary cause, insofar as it would require a bypass of the cognitive level associated with agency (Bhaskar, 1989/1998: 93). Additionally, this also implies that, from a strict ontological perspective, one should distinguish between individual action and the reproduction of social structures, insofar as one cannot derive the social structures from individual action and neither can one derive individual action from knowledge about socio-historic social structure (Bhaskar, 1989/1998: 37). This may, then, be viewed as an argument against the deductive foundation of the analytical sociological approach to mechanisms, alt-

though here mechanisms may, at least to some extent, be associated with middle range theories explaining the transformation of structural conditions into actual action – hence, the underlying logic is deductivism but the theoretical scope is not macro-social implying that the authors are not postulating, directly at least, that one can derive social structures from actual action (see, for example, Hedström and Udehn, 2011: 29). The requirement of reason does not, as already argued; imply a strict focus on actual reasons, insofar as this would lead to the impossibility of inquiry associated with the *Cartesian doubt*<sup>4</sup>, rather the focus here is on identifying possible reasons at a more general level – reasons which are not influenced by, for example, the emotional state of the acting agent (Descartes, 1637/1998: 18; Bhaskar, 1989/1998: 94).

Based on this theoretical frame, then, it follows that the problem(s) of interest/phenomena must be analyzed in a manner which allows these general mechanisms to be identified, described, and discussed. Theories closely related with a particular phenomenon of interest allow one to explain the particular phenomenon, however, these theories are also centered upon particular mechanisms which need to be explained – theories related with different strata of the phenomenon must be introduced. A general figure can now be presented:

Figure 2



<sup>4</sup> According to Descartes (1937/1998: 18), one cannot obtain real impressions, insofar as the ideas represented by the senses may as well be imaginary impression, because the real perceived may appear just as real in, for example, a dream. Thus, the only matter which can be assumed to exist is the “I” the thinking subject.

The purpose of this figure is solely to illustrate the causal chain associated with explaining the relation between a compensation policy and relational influence. The fairness of a compensation policy can be divided into the fairness of difference and the fairness of differentiation (see chapter II). The relation between difference/differentiation and acceptance can be associated with the mechanisms which enables the transformation of these results of the compensation system into acceptance. The social norms on qualitative differences, for example, can further be associated with the emotional foundation of norms and the psychological need for relatedness which combined may explain the bearing of the norms upon the agent's judgment (see chapter VIII). As argued above, the normative expectations may be associated with the horizontal and vertical relations.

The philosophical perspective presented above is often summed up in the *credo* – reality exists independent of the researcher and, at a more general level, the social agents who take part in this reality. While this assumption is in accordance with a critical realist perspective, it seems that, at least, some of the complexity inherent to the perspective is lost, insofar as the problem of reality is transformed into a problem of knowledge – however, the question is not how reality is understood in a subject-object sense, because if it is correct that understanding draws on already existing frames of interpretation, then, the question of reality becomes a social one – that is, the question becomes, how social agents are able to understand one another, and as such, understanding of the social world entails that social agents understand one another. The social world is, of course, of a different kind than the material world insofar as, for example, doors, tables, etc. exists materially independent of the perceiving social agent, although the facts of the social world can manifest themselves as just as impregnable as material facts – metaphorically – a material wall can appear just as impossible to climb as one purely erected by beliefs, norms, etc. From a pragmatic perspective, then, a distinction between a material and social world seems, at least to some extent, to be less informative. Furthermore, just as one cannot escape natural laws, such as gravity, one cannot escape social laws, in the sense that these exist independent of the acting agent, manifested in the judgments by other social agents, and activated within the acting agent as possible explanations of why these other social agents acts towards her the way they do. Social laws, in other words, manifest themselves in social practices of social agents; just as they manifest themselves in judgments, cognitive representations, and emotional to-

nality (the aesthetic dimension of reality). Social reality, then, cannot be reduced to a subjective construction, insofar as it consists of other social agents, just as it cannot be reduced to pure symbolism (a reality without ontological extension), insofar as it is manifested in social practices of interaction performed and felt by *real* agents.

### **1.3 Theoretical Delimitations**

Based on the discussion above, the theories relevant to this study can be divided into three basic groups – theories on compensation policy; theories on fairness; and theories on morality and social relations.

The theories on compensation practices within organizations should focus on issues related with monetary and non-monetary rewards, just as they should focus on design issues related with the implementation and effectiveness of these practices. Furthermore, the theories should also enable a focus on practices related with differentiation which entails a focus on the causal processes of compensation systems. In particular, the focus, here, is on gaining insight on the design issues related with compensation policies – for example, practices associated with reward management and pay structure design, and the values which these come to reflect. In that sense, the aim is not, as such, to explain, for example, the new role of Human Resource Management (HRM) in contemporary organization (see, for example, Becker and Huselid, 2006; Ulrich, Younger, and Wayne, 2008), just as the primary focus here is not whether these systems motivates, retains or attract employees; rather the focus here is solely on understanding the internal social dynamics related with the interaction between the design of such a system and the normative structures inherent to the horizontal and vertical relations in the organization. Therefore, the focus is primarily on how these practices structures or influences these social relations – horizontal as well as vertical. Additionally, the focus is limited to relational norms related with input (related with the nature of the work) and outcome (related with the monetary and non-monetary rewards associated with the work). Hence, new practices on, for example, talent management (see, for example, Lawler, 2008) may structure, at least indirectly, these relations, the focus here is on the outcome dimension of these programs – that is, how these manifests themselves in relational differences. The input dimension will be studied as how these compensation poli-

cies manifest themselves in requirements on differentiation – that is, how difference, causally follows differentiation. The relevant theories must, then, view employees as the foundation of organizational effectiveness, just as the focus on difference and differentiation can be associated with theories on performance management, though the focus will be on design issues related input (performance, differentiation, etc.) and outcome (reward design). Edward E. Lawler III's authorship based on expectancy theory enables a study of these dimensions, just as the theories which follows a similar approach to reward management as he does, for example, Patricia Zingheim and Jay Schuster (1992; 2000). Moreover, the focus here is not on management incentive systems (see, for example, Faulkender, Kadyrzhanova, Prabhala, and Senbet, 2010 for an overview of recent research), but solely on employee compensation, nor will different psychological aspects of these particular compensation systems be taken into consideration, for example, issues related with work-home balance (see, for example, White, Hill, McGovern, Mills, and Smeaton, 2003), stress (see, for example, ), or employee engagement (see, for example, Gruman and Saks, 2011).

The second and third groups of theories are related with fairness within a social context. Fairness theories are primarily related with the vertical relation – either between the organization and the employee or between the giver and a receiver. That is, fairness related with compensation is often studied solely from a vertical perspective, although some theories have viewed fairness related with the social context (see, for example, Ambrose, Arnaud, and Schminke, 2008) and fairness issues related with team-based work-relations (see, for example, Colquitt, Noe, and Jackson, 2002). Therefore, the focus has primarily been on what drives vertical demands on rewards and on which rights these demands/claims are based, for example, on performance differences (see chapter III). Extending the scope by introducing the horizontal dimension may, then, be done through a distinction between *receivers* and *non-receivers* which entails that one should take into consideration norms on receiving and not receiving – issues related with distributive justice and morality in a manner not reduced to team-based work-relations. That is, instead of asking “is my outcome fair?” a theory which introduces the horizontal dimension needs to take into consideration the two related questions, “is the compensational difference fair to me?” and “is the compensational difference fair to the other?” Hence, the introduction of the second question implies that social agents do also

hold beliefs related with the outcome of others – social preferences – without reducing the phenomenon to be primarily related with team-based work-relations.

However, issues related with justice and morality are also, at least from an intellectual historic perspective, related with rationality and social order, thus a theory which aim at explaining the horizontal dimension of fairness must also take into consideration the relation between rationality and morality. In particular the classic theories on rationality are relevant, because these do not distinguish between rationality and morality; rather they consider the two areas as intertwined in a practical dimension. The point of departure in rationality is also relevant, because it introduces the process of reasoning into the theory in a manner which enables one to distinguish between horizontal fairness based on rational and irrational reasons. Moreover, this distinction is also important due to the modern approach to economics which distinguishes between rationality and morality in manner which enables them to categorize deviations from rationality (in an economic sense) as irrational (see chapter III). In that sense, the primary focus here is on the relation between outcome differences and rational as well as irrational bases of acceptability related with judgments on the two primary issues, “fairness for me” and “fairness for you”. From this distinction it also follows that fairness, here, is primarily viewed as related with the practical dimension of organizational life, and as such it is related with, for example, psychological needs and moral judgments which imply that fairness will not be viewed as related with issues regarding trade unions. Solidarity, as it is used here, then, refers primarily to the normative bases of social life and not to some *artificial* and indirect tie generated through, for example, collective agreements. That is not say, however, that unions are not important, though it might be argued that they are of less importance here, insofar as the focus is on judgments on fairness which occurs within a social context. Furthermore, because the focus is on the relation between rationality and fairness, I will not take into consideration theories on institutional logic, insofar as these primarily focus on the relation between organizational action and institutional logics, just as rationality within this theoretical framework takes on a different meaning in the sense that these theories originally were developed in opposition to rational choice sociology (see, for example, Meyer and Rowan, 1991: 51; DiMaggio and Powell, 1991: 64; Friedland and Alford, 1991: 251). Although, the authors belonging to this perspective may be correct in their understanding of action as meaning and acting

as the externalization of socially constructed representations of particular meanings, implying thus that rationality is rationalization within a given institutional frame, this perspective, at least to some extent, seems to understate the implications of human rationality, just as their social constructivist inspired approach direct the focus towards symbolic representations and not judgment.

Even though the focus here is primarily on horizontal relations, the normative expectations may be viewed as co-produced by the normative expectations inherent to the horizontal dimension, just as it must be assumed that at least some of the general principles constitutive of vertical normative expectations also applies to the horizontal dimension, for example, reciprocity or intentionality. Therefore, theories describing the normative expectations related with the vertical dimension may be introduced as inspiration from which the basic mechanisms can be deduced. Moreover, the introduction of the vertical and horizontal normative expectations also provides a more complex understanding of the basis of judgments on fairness, hence although analytically separate; the two social relational dimensions should, in practice, be viewed as simultaneously contributing to the complexity of judgment.

The focus here is on the interaction between outcome fairness and the horizontal and the vertical dimensions, therefore I will not distinguish between distributive, procedural, and interactional fairness unless the difference makes a qualitative difference. Procedural fairness will primarily be used with reference to the evidence and foundation on which the decision to compensate actions of some employees, whereas interactional fairness primarily will be applied with a reference to the interaction between the evaluator and the employee – though I will not focus on issues related with interactional fairness, for example, the impact of authority.

The focus here is on fairness as a result of judgment which entails a focus on the formation of judgment and the factors which influences this. Basic psychological needs, such as the need to belong, are taken into consideration, just as social norms and other heuristics which may influence the formation of the judgment. In that sense, it is assumed that social agents are able to pass judgments on the two basic questions related with outcome fairness. To some extent, then, it is argued that social agents have the capacity to be reasonable in their claims. Reasonableness is viewed as the result of, for

example, morality, the need to belong, the emotional tonality of social norms, empathy, etc. Hence, reasonableness is not viewed as the result of an evolutionary process as, for example, the one described by Steven Pinker in his work “The Better Angels of our Nature” (2011). In his work, Pinker describes how violence has decreased (or at least not increased) as the outcome of an evolutionary process based on, at least, three central pillars, the creation of order/centralization of power, the civilization process, and the humanitarian revolution associated with, for example, the ending of the use of torture, decreased use of the death penalty, increasing literacy/access and spread of knowledge, and, at a more general level, higher qualities of life. Although, Pinker’s theory explicate some important and interesting conclusions, just as it seems plausible that issues related with morality, justice, etc. are contingent upon a general cultural frames of interpretation, the focus here is not to explain the genesis of morality and justice; rather the focus here is to explain how social agents formulate judgments on outcome fairness which, at least to some extent, presupposes the existence of these frames. Furthermore, closely related with this, is also the current discussion in moral psychology on nature/nurture – that is, whether, for example, morality is a result of human nature or social/cultural frames, just as it is related with a modern critique of rationalism as the foundation of moral judgment from a perspective inspired by David Hume’s understanding of morality as based on emotional representations and sensation – not reason (see chapters III and VIII). The authors belonging to this tradition explicates, for example, some important relations between social frames and emotional representations. However, the main conclusion which can be drawn from this tradition is that reason does not matter or, at least, that reason is inferior to emotions in matters related with moral judgment (see, for example, Prinz, 2012; Nichols, 2004). Here it is assumed that reason, emotions, and social norms all contribute to moral judgment, although reason is assumed to dominate insofar as judgment is assumed to be a rational process in which the agent tries to transcend the immediate situation to reach a more general conclusion, at least, manifested as the tacit claim generality accompanying the judgment in form of the belief that a generalized other given the data possessed by the individual would reach a similar conclusion. This assumption follows from the distinction between rationality and irrationality discussed above.

## **Chapter II**

### **INDIVIDUALIZATION, INDIVIDUALITY, AND THE DRIVERS OF INEQUALITY**

#### **2.1. Introduction**

In the previous chapter I discussed the philosophical and metatheoretical foundation of this thesis. The aim of this chapter is to discuss and analyze how pay a system centered upon the values, individualization and individuality may influence the relational structure of employee-colleague relations.

The two values can be viewed as logical complements insofar as individualization of the employment relation is accompanied by allowing different choices to be associated with it, for example, the choice of compensation package and the choice of benefits which can also be viewed as symbolically representing a higher degree of recognition of the individual employee's individuality.<sup>5</sup> In a similar vein, the focus on paying the person and not job in regard to the determination of base pay can be viewed as recognition of individuality which leads to an individualization of the employment relation. Pay-for-performance can be viewed as a variable pay form which allows recognition of individual differences, while simultaneously leading to a greater sense of individuality, insofar as it draws attention to the contribution of the individual to the organization's goal. Moreover, pay-for-performance also redirects the locus of control from the organization to the individual employee by allowing her a greater sense of control over her income which can be viewed as supporting this recognition of individuality.

This dual focus on individualization (focus on the individual) and recognition of individuality gives rise to several problems of securing vertical fairness in the employment relation – that is, problems related with, for example, the measurement of performance inasmuch as the goal is to secure unbiased measures, measures not biased by human error or technical inefficiency, and more generally the performance evaluation process and the principles on which this process is centered (interactional and procedur-

---

<sup>5</sup> It should be noted that individuality and individualization is here used to describe the horizontal implications of the two values which come to structure the employment relation – the employee-organization relationship – in a different manner than, for example, Ulrich Beck (2002: 89) who focus solely on the implication of these two values in relation to the spatiotemporal structuring of the relation.

al fairness) (Corby, Palmer, and Lindop, 2008: 12). Besides securing a fair process, the organization's pay policy must also secure a sense of outcome fairness which is partly vertically associated with securing balanced exchanges of input (performance) and outcome (total compensation), and partly horizontally associated with securing relative equity among employees (Milkovich, Newman, and Gerhart: 2011: 307).

These two values, individuality and individualization, can be associated with the *New Pay* trend which emerged in American compensation management literature in the 1990s as a reaction on changes in the organizations' environment due to, among other things, economic recession (Corby, Palmer, and Lindop, 2008: 7); demographic changes; loss of competitiveness due to globalization and declining organizational effectiveness; and a transition from production industry to knowledge-intensive production (Lawler, 1990: 9; Schuster and Zingheim, 1992: 4, 9; 2000: 24). New pay, in other words, emerged as a reaction on the critical situation facing organizations in the 1990s. New pay is, however, not just a compensation technology, because it introduced a more holistic approach to compensation management, strategic reward management. The strategic approach emphasized aligning the organization's goal and business strategies with its organizational design, management style, and compensation and HRM-strategies (Lawler, 1990; 1992). It was inspired by the 1960s and 1970s management literature on job-design – the job-enrichment movement – associated with authors like, among others, Frederick Herzberg (1968/1987: 12) and Douglas McGregor (1960/2006: 67), in the sense that the focus was on creating inspiring jobs which motivated individuals to participate in the organization's processes, just as this movement redirected the agency-problem from one related with human nature to one related with management and organizational design – that is, this movement can also be viewed as an attack on the assumptions on human nature which permeates, for example, Adam Smith's philosophy on rational man (see the discussion on Smith's philosophy in chapter III) and Frederick Taylor's (1911/1998: 8, 11) scientific management. Furthermore, these concerns may also be viewed as a response to Taylor's (1911/1998: 10, 14, 16, 67) rigid focus on job-descriptions and the active use of piece-rates (McGregor, 1960/2006: 71, 132; Pfeffer, 1994: 96).

New pay theorists, and especially Edward E. Lawler III (1971: 275; 1990; 1992), tried to combine the piece rate perspective with the lessons learned from job-enrichment, because as he argued, pay motivates if combined with less bureaucratically designed organizations and a higher emphasis on organizational democracy, empowerment, and involvement. The leadership-style, for example, seems to have been inspired by the 1980s theoretical focus on transformational leadership (see, for example, Burns, 1979); Japanese management style associated with theorists such as, for example, William Ouchi and Raymod Price's (1978/1993) theory on Type Z Organizations, Edward W. Deming's (1982/2000) Total Quality Management, and Tom Peters and Robert H. Waterman's (1982/2004: 75, 282) theory on leading through culture (Lawler, 1986; 1990: 11). In that sense, Lawler tried to combine some of the new theories which emerged in the 1960s to 1980s into one distinct approach to organizational management while still maintaining a focus on the significance of pay – a perspective which was contrary to the perceptions of Herzberg (1968/1987: 12), McGregor (1960/2006: 54), and Deming (1982/2000: 66) who all argued that pay, in form of contingent pay could not motivate and might, at some occasions, have the adverse effect of degrading human nature.

In that sense, new pay tried to combine a distinct view on organizations and management with a novel view on the role of compensation and HRM – a view which centered these at the core of strategic organizational management. However, because the trend was born in a period of crisis, the focus was also on how the organizations could regain a sense of control and be more effective. Hence, the theorists behind this trend focused on how compensation and HRM could be used to meet these requirements (Lawler, 1990: 5; Schuster and Zingheim, 1992: XV). Pay-for-performance was viewed as important, because it allowed the organizations to transfer some of the risk associated with compensation to the employees by making rewards contingent on performance, either individual, team or organizational (Schuster and Zingheim, 1992: 154). In addition to pay-for-performance initiatives, the organizations' compensation policies should also support organizational development, hence the compensation policy should also motivate employees to personal growth through acquiring new capabilities which could support the organizations' core competencies; just as cost-intensive bureaucratic and hierarchical goods should be removed – that is, entitlements associated with seniority

and hierarchical positions should be reduced or removed along with the entitlement inherent to job-based pay (Schuster and Zingheim, 1992: 161).

The literature on compensation management in the 2000s is characterized by an emphasis on developing the human potential and designing long-term and mutual beneficial relations between the employee and the organization. Jay Schuster and Patricia Zingheim (2000), for example, wrote about paying people right – paying them for the right things and paying them in the right manner; and Lawler (2000; 2003; 2007) wrote about mutual beneficial relations as virtuous spirals, rewarding excellence, and talent management. The emphasis is still on strategic organizational management and effective organizations, although the means proposed, here, are less focused on using compensation management as a means to transfer risk (Pfeffer, 1998: 162), and more focused on the strategic position of HRM as a tool to secure the qualitative dimension of the organization's capabilities (Prahalad and Hamel, 1990; Barney, 1991; Lawler, 1994: 13), and as such the literature is centered upon explaining how HRM is becoming more important as the need for competent employees increases, just as the distinction between micro-focused HRM strategies on human motivation and macro-focused HRM strategies on organizational performance is sought integrated through this strategic perspective (Lawler, 2007: 10; Huselid and Becker, 2011: 423). In that sense, the literature in the 2000s continued the trend from the 1990s with some minor changes in the approaches proposed.

## **2.2 Motivating the Individual to Perform**

Lawler is one of the key figures behind the popularization of performance centered organizations. Inspired by Victor Vroom's (1964/1995) expectancy theory, Lawler proposed, together with Lyle Porter (1968), a motivational theory, which he later developed into a theory on how organizations should be constructed to secure highly efficient work organizations, what he called the High Involvement Organization. His work on high involvement organizations can be viewed as a combination of a management approach, organizational design, and HRM, insofar as he demonstrates how these must be combined in order to secure the foundation of organizational performance. In 1990 he published an influential book on how pay could be used strategically to secure organiza-

tional performance by aligning the organization's pay system and more generally its HRM-practices with the organization's goal. This book inspired the today commonly held perspective that pay should be used strategically to support the organization's goals, but also to secure organizational development in the sense that the organization continuous development should be secured through its employees continuous development (see for example, Milkovich, Newman, and Gerhart, 2011 or Armstrong and Brown, 2009).

In order to understand the theories which will be discussed in the following, it seems appropriate to start with motivation, insofar as it seems to be at the core of this perspective, thus I will in the following section introduce some core assumptions on work motivation found in expectancy theory. Expectancy theory provides a process theoretical view on behavior in the sense that it focus on the relation between individual perception and the environment, the individual perception of the environment generates the foundation of subjective beliefs/subjective expectations which, subsequently, come to be the foundation of the interaction between the individual and the environment (Fugde and Schlacter, 1999: 296; Isaac, Zerbe, and Pitt, 2001: 214). Furthermore, expectancy theory focuses on how extrinsic motivators cause behavior (Leonard, Beauvais, and Scholl, 1999: 972). Later in this thesis, I will introduce different motivational theories which are critical towards the implementation of, for example, performance-related-pay (see chapter IV).

According to Lawler (2000: 77; 2003: 38, 39, 51), the human agent is mostly rational, and as such the agent is: (a) future oriented; (b) utility maximizing; and (c) engage in instrumental reasoning regarding how best to satisfy her needs and reach her goals (Shamir, 1990: 316). Assumption (a) implies that the agent is motivated to act in ways she believe will lead to valued rewards, that is, the agent is motivated by the promise of a reward in the sense that motivation is based on the anticipated future satisfaction which is believed to follow from obtaining some reward – the performance-outcome expectancy (Lawler, 2000: 79; 2003: 39, 51). Assumption (b) implies that the agent: on the one hand, chooses the reward which best satisfies her active need (Lawler (2003: 40) is here inspired by Abraham Maslow's (1943) need hierarchy); and, on the other hand, prefers more to less (Lawler, 2000: 71; 2003: 40, 45). Combining assump-

tion (a) with (b) implies that the agent must perceive the promised reward as meaningful or important if it is to have a motivational impact (Lawler, 2000: 70; 2003: 40). These need-deficits felt by the agent, which lend valence (meaning or importance) to a reward, can only be satisfied from the outside – through extrinsic rewards – albeit Lawler (2000: 78; 2003: 41) does not neglect intrinsic rewards – rewards which stems from the inside or from the agent herself, such as, for example, self-esteem or personal growth. Furthermore, besides being influenced by lower and higher order needs, the valence of the reward is also influenced by the environment within which the agent finds herself, that is, the agent may be influenced by other agents opinions or she may be inspired by seeing other agents succeed in obtaining certain rewards – this area is also particularly important with regard symbolic rewards (Lawler, 2000:70; 2003: 41). Assumption (c) implies that the agent, given her subjective expectations, chooses the action she believes to be most effective in bringing about the desired reward, just as it also seems to imply that the agent is egoistic in nature, insofar as it is *her* needs and goals which she tries to bring about (Lawler, 2003: 38; Shamir, 1990: 316). Moreover, assumption (c) also implies that the agent hold the meta-belief that it is within her power to bring the desired reward about ( – it is, in other words, a matter of perceived outward directed causality, insofar as the agent must believe that her actions actually have an impact or, at least, that she is able to control or manipulate (in a non-negative sense) her environment, what Lawler (2000: 77; 2003: 48) calls establishing a clear *line of influence*. From this assumption it also follows that the agent must perceive the action as being within her space of possible actions – that is, she must believe that she capable of bringing the state about in the sense that she must believe that she possess the necessary abilities, skills, competences, etc. (Lawler, 2003: 55, 57). Lawler (2003:56) distinguishes between *fundamental abilities* inherent to the individuals’ predispositions; and *acquired dispositions* which are primarily obtained through education, training, etc., while it may be problematic to expand the former, the latter can be strategically targeted through HRM-strategies. This may now be formalized into a simple equation:

$$Behavior = f(Motivation; Ability; Environment).$$

Or in Lawler’s (2003: 38) formulation:

$$Performance = f(Motivation; Ability)$$

Lawler's (2003) formulation can be viewed as a reformulation of the formalization of behavior as a function of motivation and ability. However, it should be noted that Lawler (2003) does not include the environment in his equation, though he explicitly discusses it. One reason may be that the environment cannot be strategically controlled in a similar manner as motivation or ability, a speculation which seems to be supported by the examples he provides on the symbolic dimension of rewards (Lawler, 2000: 73); although one might argue that environmental factors such as organizational design, job design, organizational culture, and management are strategically important insofar as they might have averse consequence to the performance of the individual if not aligned with initiatives targeting motivation (reward strategies) and ability (human development strategies) (Lawler, 1990: 29, 31). Within the context of this section, then, recognition of the individual is associated with motivation and individuality is associated with both motivation and abilities insofar as the driver behind individuality besides a focus on ability also includes a recognition of heterogeneity, that is, social agents are different from one another and in order to have an impact on performance, reward systems must support these differences (Lawler, 2000: 74). So, individuality is also linked to reward satisfaction.

In addition, pay satisfaction is also related to outcome expectations related with a hypothetical standard held by the employee as being an equitable exchange, and relative rational outcome expectations formed by horizontal (internal and external to the organization) social comparison, and on some occasions the employee might also hold some relative irrational outcome expectations formed by upward or downward social comparison (Lawler, 2000: 80, 81; 2003: 45, 46). Outcome satisfaction is also influenced by concerns about procedural fairness. Procedural fairness may be secured through: (a) openness about the process of decision-making (Lawler, 2000: 82; 2003: 47); (b) including the right and trusted people (Lawler, 2000: 82; van den Bos, 2001: 73); (c) state the criteria on which the assessment is founded clearly and in advance (Lawler, 2000: 82; 2003: 47); (d) allow the employees to participate in the evaluation process (Lawler, 2000: 82; 2003: 47; van den Bos, 2001: 69); and (e) allow the employees to appeal the results of the evaluation process (Lawler, 2000: 82; 2003: 47).

The individual employee's fairness expectations are, then, primarily based on rationally formed expectations. The fair outcome is that which an employee in a similar job either within the same or in another organization obtains. The fair outcome is, in other words, that which a rational agent could expect to obtain. The rational is also, according to this, the reasonable. The fair outcome is, as such, that which an employee based on her rationally formulated expectations finds reasonable.

## **2.3 Individualization**

After this short introduction to Lawler's expectancy theory, I now turn to the first of the two values inherent to contemporary approaches to compensation management, individualization. Individualization is, here, associated with two trends: (a) base-pay should be based paying the person for the knowledge, competencies, and skills she holds; and (b) the employees' individuality and heterogeneity should be recognized through individualization of the employment relation.

### **2.3.1 Pay the Individual**

Pay the individual and not the job seems to be a persistent slogan in current approaches to compensation management and have been since the 1990s (Pfeffer, 1994: 45). From a historic perspective, paying the individual along with the greater use of variable reward components can be viewed as a *rebellion* against job evaluation (Lawler, 1986b; 1990: chapter 8). Job evaluation can be understood as a careful and elaborate process of pricing each job or groups of jobs (job families) in order to create an internal hierarchy of jobs based on the assessment of elaborate descriptions of job requirements (Ingster, 2008; Milkovich, Newman, and Gerhart, 2011: 139). In that sense, the job is priced regardless of the individual performing the job – that is, it is priced regardless of the skills, knowledge and competencies possessed by the individual employee, thus the pay associated with the job reflects the internal hierarchy and, at least to some extent, the job's external market value (Milkovich, Newman, and Gerhart, 2011: 239).<sup>6</sup> From a fairness perspective, job-based pay employs an *objective* approach, insofar as the jobs

---

<sup>6</sup> As Armstrong and Brown (2009: 121) argues, in Europe the focus is still very much on securing internal market value, whereas the American seems to focus on securing external market value.

are priced in accordance with the horizontal and vertical pay structure, and as such regardless of the employee holding the job, hence the system allows pay to be neutral or insensitive to gender, ethnicity, social class, etc. (Armstrong, Cummins, Hastings, and Wood, 2003: 51). In that sense, job-based pay can be viewed, from a historical perspective, as an important technological development, along with bureaucracy, which enabled a modern perspective on the employment relation to emerge, insofar as it substituted loose and often ad-hoc relations with stable ones, along with the introduction of the employment contract, based on fair rules (Mahoney, 1989: 338; Ingster, 2008: 96).

Job-based pay can be viewed as being based on the fairness principle, equal pay for equal work – hence; it is centered upon the recognition of equality among individuals and job-based equity (Armstrong, Cummins, Hastings, and Wood, 2003: 57, 58). Taylor's (1911/1998) emphasis on job-descriptions and his emphasis on bureaucratic management – the manager who directs and control the work of the employee – can be viewed as having inspired the job-based approach, albeit he emphasized the use of piece-rates as an important means to overcome agency-problems within the organization. The job-based approach's emphasis on job-descriptions was highly criticized by Lawler (1990: 140, 141, 142) because it reduced the employment relation, and as such the relations; manager-employee and employee-organization, to a matter of the job – in a sense, then, it *depersonalized* the relation, and symbolically it communicated to the employees that they did not possess any value in themselves – only the job they held did, just as it, at least implicitly, emphasized what not to do by emphasizing what to do. Furthermore, because the job-based approach is highly focused on horizontal equality of jobs, it may foster a focus on vertical careers as being the *only* way to a significant change in income insofar as the employee would only obtain a higher pay by changing her job (moving up in the hierarchy), and as such it may put too much emphasis on promotion and less focus on personal development and horizontal careers which is important to secure a flexible and highly adaptable organization (Lawler, 1990: 147, 149; Pfeffer, 1995: 47). This should also be viewed in relation to another feature of the system – the creation of entitlements, because the employee holding a particular job might come to view the pay and benefits associated with the job as an entitlement inherent to the job, and as such the pay may not be associated with something which is the result of performance, but rather as a property right which cannot easily be changed or removed,

hence it might induce resistance to change if these would be perceived to influence these perceived entitlements. As I will argue in the section on pay-for-performance, the introduction of pay-for-performance can be viewed as a *rebellion* against and as a disruption of this certainty effect which follows as a consequence of certain aspects being perceived as entitlements. In chapter VII of this thesis I will further discuss the potential implications of this disruption of entitlement insofar as entitlements are more than *mere* expectations – they are perceived as claims, supported by the individuals' cognitive structuring and social norms, implicating that changing these might be viewed as a violation of rights anchored in the cognitive and social structuring of the employment relation. This, of course, also has implications for the horizontal dimension, employee-coworker, which I will further discuss in the section on pay-for-performance.

The vertical dimension of the job-based system, the job-hierarchy, at least to some extent, mimics the positional hierarchy of the organization, and as such higher pay and better benefits are associated with higher level positions. In other words, the combined effect of the compensation system and the positional hierarchy gives rise to an organizational conception of social classes (Baron and Pfeffer, 1994). That is, by pricing jobs based on job-descriptions, the organization also creates social classes based on the opportunities, income levels, and benefits associated with these different jobs (Baron and Pfeffer, 1994: 191). Horizontally, these different job-descriptions creates a sense of belongingness through categorization – that is, employees belonging to the same job category might come to view themselves as similar to those belonging to the same category, while those belonging to other job-categories are perceived as belonging to another group, just as positional hierarchy delimits group-membership vertically (Baron and Pfeffer, 1994: 196). These social groups which emerges as a result of job-categorization can be viewed as being based on a sense of common/shared fate – insofar as the group members belonging to another horizontal job-category will not only hold other jobs belonging to another category, they might also have different compensation packages attached to their jobs, hence by assigning different employees to these different jobs, and by assigning different pay levels to different jobs, the organizations are creating a nesting place for symbolic differentiation – social classes based on the relation employee-job (Baron and Pfeffer, 1994: 194). Inequality, then, is not just materially manifested in different pay levels associated with different jobs; it is also symbolical in the sense that,

for example, jobs requiring unskilled labor are found at the bottom of a job hierarchy, and these jobs being associated with low pay (Baron and Pfeffer, 1994: 197). That is, by creating such pay hierarchies based on job-categorizations, the organizations are, in other words, structuring, socially as well as cognitive, group membership and group identification (Baron and Pfeffer, 1994: 192). Furthermore, this entails that if few objective differences exist, intergroup comparison of employment relations are more valid, thus lower intergroup inequality can be justified, just as interdependency between jobs may also reduce intergroup inequality (Baron and Pfeffer, 1994: 200; Levine, 1990: 240).<sup>7</sup>

James Baron and Jeffrey Pfeffer (1994) in their discussion of organizational inequality focus primarily on the interaction between job-categorization and compensation as creating the symbolic and monetary basis of inequality, however, one might, based on this framework, ask what happens when intragroup differentiation is introduced, that is, when individual differentiation is emphasized – more specifically one might ask, how this influences perceptions of fairness because while job-categorizations remain they lose significance as points of social and cognitive lines of demarcation, just as one might ask how this will influence intragroup relations – this will be the primary focus in the following chapters of this thesis.

Person-based structures was emphasized as an approach which would not only solve the problems associated with job-based structures, it would also generate the foundation for a new perception of compensation as the driver of personal development, insofar as organizational development is supported by the development of its employees (Lawler, 1990: 23; Lawler and Ledford, 1992). Compared with job-based pay, person-based pay compensates the individual regardless of the job she holds, and as such it was viewed as a step towards personalizing (and individualizing) the employment relation, just as it was viewed as a step towards evaluations based on the employees' external market value (Lawler, 1990: 153; 2000: 127; Schuster and Zingheim, 1992: 95; Ledford, Heneman, and Salimäki, 2008:144). Person-based pay can, then, be viewed as a

---

<sup>7</sup> Tournament theory proposes the alternative view that relative differences need to be high in order to motivate insofar as high relative monetary differences also generate symbolic differences important in distinguishing between winners and not winners or average and better performers (Connelly, Tihanyi, Crook, and Gangloff, 2013).

pay system design which rewards the individual employee for the skills, knowledge and competencies she possesses – her capabilities – and as such it can be viewed as an indirect kind of performance-related-pay because it links performance with its foundation, the individual's capabilities (Ledford, Heneman, and Salimäki, 2008:144). According to Lawler (1990:155, 156, 161; 2000: 128), Pfeffer (1994: 46) and Schuster and Zingheim (1992: 97; 2000: 98) person-based pay is a better fit with the new flat organization with a focus on horizontal careers than job-based pay was, because it emphasize personal development, multiskilling, and in-depth knowledge which support organizational flexibility and adaptability. From a fairness perspective, person-based gives rise to considerations regarding internal and external pricing of the employee's capabilities insofar as these should be fairly priced internally and externally, just as these opportunities of personal development should be offered in a manner which generates equal opportunity (Schuster and Zingheim, 1992: 97; Milkovich, Newman, and Gerhart, 2011: 175).

Compared to the job-based approach, the person-based approach allows the organization to reward the employee for qualitative differences related with the employee's input and personal development, just as it is based on the idea that the employee through personal development develops her job. The job-based approach presuppose that the individual already possesses the necessary skills, competencies, and level of knowledge to hold the job, the person-based approach, on the hand, is not build on this assumption, but views this as a continuous process of adjustment between the individual employee's capabilities and the organization's needs. Hence, the employment relation cannot just be viewed as a continuous negotiation of what needs to be done – the distribution of tasks and securement of effort – but also a continuous negotiation of which capabilities the organization needs its employees to possess (Schuster and Zingheim, 1992: 96; 2000: 97; Lawler, 2000: 128). First of all, person-based pay is to be viewed as part of the total compensation offered to the employee, insofar as it is related with equitability and personal development (Schuster and Zingheim, 2000: 13). However, it may also be viewed as a broadening of the authority relation inherent to the employee-organization relation, and as such one should expect this to be governed by the same rules as the distribution of tasks – that is, these demands on further education and personal development must take place within a *zone of acceptance* (Barnard, 1938/1968: 168; Simon, 1951: 294, 295; 1945/1997: 10; Marsden, 17, 21; 2000: 324; 2004b: 663),

and by attaching pay to personal development, the organization can *renegotiate* the scope of this.

Paying the individual and not the job, then, allows base pay to reflect qualitative differences associated with the individual employee's capabilities, and as such it reflects the individual employee's ongoing value, just as it allows a long-term perspective on qualitative performance changes (Schuster and Zingheim, 2000: 97, 109). Promotion and pay raise should be based on changes in these capabilities and not performance or seniority (Lawler, 2003: 187). Paying the individual, in other words, emphasizes mutual beneficial and long term relations between the employee and the organization – relations based on mutual development. As such, this trend seems to be related with the recognition of the importance of human resources as the strategic foundation of the organization, just as it can be viewed as the necessary foundation of securing high-involvement principles of organizing and empowerment, because, as also argued above, motivation is only one part of the driver of performance and without being in possession of the *right* capabilities, motivation does not lead to qualitatively better performance – only quantitatively higher levels of effort, just as it might discourage an employee from pursuing a particular goal if she does not believe herself to possess the necessary capabilities.

Based on Baron and Pfeffer's (1994) perspective on organizational inequality, one may argue that the rise of this trend in compensation management has shifted the lines of demarcation from being based on job-descriptions to new ones based on individual differences associated with the qualitative dimension of the employee's input. In the classic bureaucracy higher levels of education and training was often associated with higher levels in the positional hierarchy, whereas with the introduction of paying-the-person initiatives, higher levels of capabilities are not associated with positional differentiation, hence two employees, for example, who hold similar jobs and positions within the same organization may obtain different levels of income due to the level of recognized capabilities they possess, and in the section below I will add that these employees may also receive differentiated compensation packages. The *cause* of inequality, then, is not *external* and *objective* differences; rather it is individual differences in recognized capabilities. Furthermore, the introduction of paying-the-person allows inequal-

ity to enter within-group relations insofar as these are not based on horizontal and vertical intergroup lines of demarcation as, for example, job-based pay are. Here it might be expected that fairness concerns are important, insofar as these qualitative differences must be justifiable, just as equal opportunities may matter, because the employee who refrained from accepting an offer on, for example, in-job-training cannot perceive it as unfair, if another employee who accepted the offer obtains a pay raise – or can she? I return to this question in chapter X.

### **2.3.2 Individualization of the Employment Relation**

The trend towards a greater individualization of the employment relation seems to be driven by the question “if people are different, why should they be offered the same reward package?”, just as it is driven by a libertarian egalitarian assumption on differences stating that these should reflect individual differences and not positional differences, or differences in jobs held (Lawler, 1990: 217; 2000: 54; 2007: 78; 2011: 303). In other words, the compensation system should be administered in a way which allows individual differences and preferences to be taken into consideration. From the organizations’ perspective, the driver behind individualization seems to be that they need to attract and retain highly competent employees (Lawler, 2007: 79). Moreover, the trend towards allowing individual choice can be viewed as a *rebellion* against the paternalistic value foundation inherent to standardized compensation packages, insofar as the organization by allowing the employee to design her private compensation package, symbolically, communicates to the employee that it has faith in her ability to make an informed choice on what is best for her (Lawler, 1990: 217).

This can be viewed in relation to expectancy theory, insofar as the compensation in order to motivate the employee should satisfy an active need (Lawler, 2011: 304). In relation to the design of compensation systems, the introduction of individualized compensation packages need to allow the a employee choice with regard to both delivery and content – the employee, for example, must be allowed to choose between money and benefits, just as she must be allowed to choose between which benefits she desires (Lawler, 2007: 79; 2011: 305).

The recognition of heterogeneity and the implementation of these in the employment relation invoke some important problems with regard to outcome fairness, because the employees cannot compare their total compensation packages in the same manner, or at least with the same evaluative validity, as they would have been able to, had it *only* been wages/salary. This should also be viewed in relation with the fact that too much individualization can reduce the employees' sense of community and a common fate (Lawler, 2003: 77-78). Furthermore, as Lawler (2011: 303) argues, treating employees differently may give rise to fairness issues insofar as an employee who is less satisfied with her self-composed compensation package may believe her more satisfied colleague is getting a better deal. What Lawler is here arguing is that solidarity, belongingness – the sense of a common fate – cannot be reduced to quantitative matters as the conventional perspective seems to do (see, for example, Cook and Hegtvad, 1983: 223). To overcome this problem, Lawler (2007: 81) proposes that the organization should, at the point of hiring, explain to the potential employee that the employment contract is individually negotiated, just as they should emphasize that it reflects external market realities. This seems to suggest that inequity is not associated with securing a balance between input and outcome based on internal equity; rather securing equity becomes a matter of external equity – what the employee could get in similar job at another organization. Just as it becomes a matter of designing the compensation package in a manner which motivates the individual employee, because it allows her to design it in a manner which best satisfies her needs. Furthermore, allowing the employee to design her compensation package, within certain limits, also, at least to some extent, secures that the rewards received are meaningful to the employee, just as it secures that it is something which the employee recognizes as part of the exchange (outcome).

Additionally, the *menu* from which the employee should be allowed to choose from should not be designed in a manner which allows privileges to be attached to certain positional levels – in other words, the benefits from which the employees should be able to choose should be egalitarian in nature, providing all employees, regardless of position in hierarchy, the same choices (Pfeffer, 1994: 49; Lawler, 1990: 212; 2000: 52). Symbolic egalitarianism can be viewed as an important means to promote a sense of common fate across different horizontal jobs which enables information sharing between different jobs and departments within the organization, just as it supports infor-

mation sharing between different hierarchical positions within the organization. However, whereas symbolic egalitarianism seems to have as its aim to reduce the sense of structural differentiation, individualizing employment relations may be viewed as substituting structural differentiation with individual differentiation. Stated differently, symbolic egalitarianism reduces or removes vertical symbolic privileges, individualization produces horizontal symbolic differences based on recognition of individuality. Moreover, whereas the differentiation based on vertical structures may have status as a social norm insofar as it may be viewed as almost inherent to the employment relation, differentiation based on recognition of individuality may be perceived as a threat to within-group solidarity. That is, if Baron and Pfeffer's (1994) claim that within-group relations are based on similarity of conditions is correct, allowing individuality may reduce this sense of perceived similarity which cognitively structures social affiliation. Moreover, this reduced sense of similarity may also be viewed as a threat to the need to belong. I return to this in chapter VIII.

Recognizing individuality by allowing choice can, along with person-based pay, be viewed as two trends which allow individuality to enter the employment relation as a foundational value. This can lead to two different kinds of horizontal within-group inequality: (1) within-group inequality based on recognition of the individuals' ongoing value in terms of external market value (Lawler, 2000: 47) and capabilities can lead to different pay levels emerging for similar jobs within the same organization; and (2) within-group inequality based on recognition of individuality can lead to individually differentiated compensation packages.

### **3.3 Recognizing the Individual**

In the previous section I discussed two kinds of individualization. In this section, I turn to the second value inherent to contemporary compensation systems, individuality. Individuality can be viewed in relation to pay-for-performance initiatives in particular and more generally in relation to employee performance management, which both demands the individual to differentiate herself.

#### **3.3.1 Pay-for-Performance**

Pay-for-performance and performance appraisals combined generate the foundation of performance management which can be viewed as one of the most important elements in contemporary approaches to compensation management. Performance management, then, has a monetary and a non-monetary dimension, the focus of the non-monetary dimension is personal growth, development, career possibilities, and goal-setting (Lawler, 2003: 163, 175; 2007: 102; Wilson and Malanowski, 2008: 448-449). Pay-for-performance is part of the individual employee's total compensation package and covers the variable dimension of the employee's total pay, whereas person-based pay is related with the fixed element, base pay. The introduction of pay-for-performance plans can be viewed as the end of the entitlement associated with fixed pay levels insofar as the employee will have to earn it at every period (Schuster and Zingheim, 2000: 109). The driver of income here is individual performance implying that the better performer may be able to earn an income above the market level, while a low performer might earn an income below the market rate (Lawler, 2000: 150). Pay-for-performance might, then, have a sorting effect insofar as low performers sort themselves out.

Pay-for-performance as a variable element of total pay was, originally, designed to overcome the problems associated with merit pay, because in merit pay positive performance evaluations lead to increases in base pay (Lawler, 1990: 72; Pfeffer, 1998: 211). According to Lawler (2003: 185), performance-pay should then be based on lump-sum bonuses, insofar as this creates a clearer line of influence.

By making a part of the employee's total pay contingent upon reaching some pre-established goal or standard enables the organization to, more or less, directly influence the employee's choice of performance level (Lawler, 2007: 110). The introduction of performance-related-pay installs a clear line of influence, insofar as it enables the employee to influence her income and tells how she should do it (Lawler, 2007: 102). In that sense, pay-for-performance enables the individual employee to influence her own income level, and as such the employee becomes responsible for her own income – at least the variable part – implying that the employee gain a sense of causal power over her income. This sense of causal power is contingent upon the factors mentioned in section 2.2, just as it is influenced by the employee's understanding of the link between

performance and rewards; hence employees should be informed about how the system functions.

Pay-for-performance can be, but need not be, based on individual performance, in some cases it might be more appropriate to make the variable element contingent upon team-performance or organizational performance depending upon the nature of the work and the line of sight the organization wishes to install, insofar as making the employee's pay contingent upon organizational performance may install a longer line of sight focusing on aligning the individual employee's goals with organizations' goals (Schuster and Zingheim, 2000: 148). Pay-for-performance can, in other words, be viewed as a vehicle for communicating what needs to be done, which kind of behavior is appreciated, and what is important to the organization (Schuster and Zingheim, 2000: 143). While longer lines of sight can be good at aligning the individual employee's behavior with that of the group or the organization there might be a trade-off with regard to the employee's sense causal power, insofar as the reward is not only contingent upon her behavior, but the behavior of the other employees involved and other external factors outside the control of the employee. Just as a trade-off might be identified in regard to the symbolic communication of the pay-for-performance initiative, because the symbolic communication does not solely cover the appraisal dimension related with celebrating performance, it also covers the symbolic implementation of features of symbolic distinctions, insofar as it allows the organization to distinguish between high and low performers and pay-for-performance initiatives based on team or organizational performance will typically pay an equal reward to everyone regardless of individual contribution, and as such high and low performers receives an equal amount (Lawler, 2000: 221).

The symbolic dimension of performance management and especially pay-for-performance initiatives is contingent upon two effects: the *celebration-effect* and the *recognition-effect* (Schuster and Zingheim, 2000: 146, 189). While recognition can be communicated through non-public vehicles of communication, for example through bonus payout, celebration is a public affair. In some cases public celebration of performance may even magnify the recognition-effect of a bonus (Schuster and Zingheim, 2000: 195). Celebration can take on many forms, from actual celebrations to more subtle, but still public forms, such as t-shirts or mugs with the text "Be like X" or "Per-

former of the year” (Schuster and Zingheim, 2000: 193). The organization can through celebration show its recognition of the individual, just as it can produce “good examples” which might enhance the effect of the symbolic distinction dimension. While this might have the effect of singling out high performers, the organization might also use more subtle ways of publicly communicating about pay, for example, the organization might publish a list on rewards distributed, so the employees can see how they did themselves and whether they belong to the top or bottom, the important aspect here, however, is not the monetary dimension, but the symbolic dimension – that is, the employees should not compete over being the highest paid, but being the best performer.

Making results public and being open about the organization’s pay policy has two important dimensions: on the one hand, being open about the organization’s pay policy forces the organization to be more precise and fair in its evaluation processes, just as it commits the organization to its pay policy (Lawler, 2003: 196); and, on the other hand, it communicates to the employees that the organization perceives them as rational and sensible agents who are able to, if provided the correct kind and amount of information, produce rational judgments on the fairness of the organization’s processes, but also on the fairness of their own pay compared to that of another employee (Lawler, 2003: 196, 198). According to this view, then, openness enables the employees to make informed and rational decisions about their own career, just as it enables them to make informed and rational judgments on the fairness of the organization’s pay policy. This approach seems to be centered upon the assumption that employees are not just rational agents, but also reasonable agents who accept justifiable differences and differentiation if allowed to assess the foundation of the decision (Lawler, 2003: 197). Pay secrecy, in other words, hinders the employees in making rational and reasonable judgments, insofar as it may encourage the wrong type of beliefs and invoke irrational emotions such as envy. Against this perspective it might be argued that openness may confront the employee with her own position in the reward hierarchy, and as such she may be confronted with the fact that she is not doing as well as she thought which may lead to dissatisfaction (Pfeffer and Langdon, 1993: 386).

In the section on motivation it was identified that the reasonable outcome is based on rationally formed expectations. From what has just been argued it seems that the

reasonable agent will accept the distribution of outcome if she, based on a rationally founded judgment, finds the process fair. Pfeffer and Langdon (1993) argued that the opposite might be the case, that certainty of one's position within a pay hierarchy is worse than uncertainty. From another perspective, one might argue that the rational agent portrait here holds no expectations which are not rationally founded – based on correct information – and processed. Fairness, in this perspective, becomes a matter of judgment – the formation of judgment – but it also, as argued by McGregor (1960/2006: 126) and Nicholas Rescher (2002: 4), becomes a matter of understanding acceptance. In addition to this one may ask, what is the relation between rationality and acceptance, and does acceptance of one's own pay as fair lead to acceptance of one's place in the pay hierarchy? That is, do acceptance of one's own pay as fair lead to acceptance of other agents' outcome as fair, and do fairness exclude social emotions such as envy, guilt, or pity from being invoked? In other words, is that which is considered as fair vertically with regard to equity also horizontally fair? Furthermore, this also leads to a question regarding the very foundations of the expectations held by the employee, because do these solely consist of rationally founded expectations – do agents not also hold cognitive normative expectations, social normative expectations, and moral normative expectations?

#### **2.4 Individual Differentiation, Social Norms, and the Need to Belong**

Another important question might also be the type of inequality which pay-for-performance creates, because it produces temporary or non-constant inequality caused by performance increases. This implies that pay-for-performance may produce a higher degree of pay dispersion within an organization (and more generally on the labor market) – this effect is also supported by empirical data (Lemieux, MacLeod, Parent, 2009: 37; Barth, Bratsberg, Hægeland, and Raaum, 2012: 346). This effect, however, may not solely be associated with an increase in the use of performance pay, insofar as the introduction of person-based structures might also increase pay inequality (Lemieux, MacLeod, Parent, 2009: 44). Hence, the inequality effect of pay-for-performance may be problematic to differentiate, statistically, from the effect of person-based pay. This is, of course, also in accordance with Lawler's (2003) equation on performance as being a

combined product of motivation and ability inasmuch as performance has more than just a quantitative dimension it also has a qualitative dimension, therefore the combined effect of initiatives on individualization and individuality might be a higher degree of within-firm inequality. Besides monetary inequality, pay-for-performance may also lead to symbolic inequality associated with the symbolic distinction associated with distinguishing between high and low performers. In that sense, pay-for-performance add to the inequality dimension associated with difference, just as it adds a new dimension, symbolic inequality, which enables the exchange of outcome into social symbols such as low and high performers. Furthermore, it also adds another driver of inequality, performance. Performance is different from person-based induced inequality because it is not only associated with differentiation in regard to the employee's individual choices regarding career and education, but also adds another dimension, differentiation through effort, at least in regard to the employee's determinateness in reaching her performance goal. That is, in a perfect world with complete information and fair assessments, the inequality caused by the variable component of total pay would be solely driven by performance differences. This implies, then, that the individual employee rated the best performer has willingly differentiated herself from her colleagues, just as she might have differentiated herself through education or in-job-training. Total compensation is, as such, based on the assumption that employees desire to differentiate themselves from one another and to be recognized when they do. However, whereas pay differences can be viewed as exogenous to the employee-colleague relation, differentiation cannot insofar as this, logically, is something done by the individual when she decide to act in a way which differentiate herself from colleagues. This self-differentiation may be magnified if the organization supports this. The individual who chooses to differentiate herself from her colleagues by pursuing some opportunity which primarily is to her benefit might, however, be considered as antisocial, not because she pursues a selfish goal; but because she acts in a way which differentiates herself from her colleagues.

This distinction between exogenous pay differences and self-differentiation may also be problematic, because whereas pay differences activate cognitive schemes associated with, for example, distributional fairness these are still exogenous to the employee-colleague relation; self-difference with regard to performance, on the other hand, activates cognitive schemes associated with, for example, workmen's pride, hence as

the following examples also demonstrate it might be easier to accept exogenously induced pay differences than self-differentiation through performance.

Consider, for example, the interaction between norms on relational differences when invoked in a situation framed by norms on outcome differences, and when invoked in a situation framed by norms on input differences. Sverre Lysgaard in his work *Arbeiderkollektivet* [The Workers' Collective] (1961/2010) reports a study from the 1950s of the workers at the organization, *M. Peterson & Søn A/S Moss Cellulose* in Norway. Lysgaard (1961/2010) observed how judgments on the fairness of pay were based upon a norm set which allowed high achievers to obtain a wage reflecting this. However, whereas high achievers did not violate a relational norm on pay fairness, they violated a relational norm on input fairness, in the sense that they differentiated themselves, by being high achievers, from the group of workers, and whereas outcome differentiation may be considered as fair, input differentiation was not. Outcome differentiation, then, becomes viewed as an act of favoritism, because as it is argued in one of the interview quotations: "*The management wants difference – the best should be the highest paid. But pay should reflect the average not the best.*" (Lysgaard, 1961/2010: 108. *Own translation from the original text*<sup>8</sup>). Here, an outcome differentiation is understood, by the respondent, as an act of favoritism even if the foundation of this is the superior skills of the employee being rewarded by the management. A pay difference takes on a particular symbolic meaning in the sense that it comes to reflect a symbolic act of differentiation – generating differences between equals – emphasizing the effort of one over the collective effort of the group, and as such reduces the pride and spirit of collective accomplishment (Lysgaard, 1961/2010: 107). The employee who actively, intentionally, attracts the attention of the management is also violating the sense of collectiveness in the sense that she tries to differentiate herself, in the eyes of the management, from her peers, and this behavior is symbolically categorized as being a *bad colleague* (Lysgaard, 1961/2010: 109; 111). Being a *good colleague*, on the opposite hand, means showing solidarity with one's peers and acts of displaced loyalty becomes categorized as acts of non-solidarity – in that sense, solidarity is mutually exclusively because one can either be loyal to the management or loyal to one's colleagues. In other

---

<sup>8</sup> Original quotation: "*Ledelsen ønsker forskjell – de dyktigste skal tjene mest. Men lønnen må stipuleres etter den gjennomsnittlige, ikke etter den dyktigste.*" (Lysgaard, 1961/2010: 108)

words, the workers collectively stands *united* towards a *common* enemy, the management (Lysgaard, 1961/2010: 110). If an employee works harder than what is prescribed by the governing effort norm, she is, intentionally, differentiating herself from the other workers in order to make herself appear better than her peers, and as such she is being egoistic and showing her loyalty towards the management instead of her colleagues (Lysgaard, 1961/2010: 111).

Insofar as the employee is acting in a way which is perceived as non-solidarity by her colleagues, she is perceived by these as an ally of the management and is likely to be ostracized by her colleagues from the group (Lysgaard, 1961/2010: 111; 113). This should, of course, not be interpreted as a legitimization of disobedience, because the employees should follow the rules of management and do what is expected from them in terms of their job, however beyond this formal demand, emerging from the contractual agreement between the employee and the organization, the employee should not be too willing and eager to take on, for example, extra work, because while the former is governed by the relation between the employee and the organization, the latter is governed by the norms of the employee-colleague relation – the worker, therefore, cannot just take on extra work without consulting the rules prescribed by the social norms (Lysgaard, 1961/2010: 113). An employee who is too willing and eager to take on extra work imputes competition between the employees which reduces the sense of a collective spirit uniting the workers, thus extra work must be distributed equally between the employees (Lysgaard, 1961/2010: 115). Additionally, this sense of a collective spirit may also be reduced by secrets – that is, if an employee, for example, decides not to tell her colleagues about how much she has earned, her colleagues might interpret this as a signal of displaced loyalty, that the employee has obtained something extra (Lysgaard, 1961/2010: 112).

Another example of how situations may be complicated by the interaction of different norm sets governing different actions of organizational life is found in the Hawthorne studies of the Bank Wiring Room. The incentive system was based on group piece work implying that the only way the earnings of all the employees within the particular department could increase was through a collective increase in total output (Roethlisberger and Dickson, 1939/1975: 409-410). The system was designed to max-

imize the production of the single employee and the group (Roethlisberger and Dickson, 1939/1975: 416). The pay of the single employee was determined as follows: The entire department was treated as one economic unit to which a fixed amount was distributed based on the department's total production and these total earnings was subtracted from the total *Daywork Value* of all employees (sum total of all employees': hourly rate<sup>9</sup> x hours worked) to find the *Excess piece-rate earning*, which was, then, divided by the Daywork Value to obtain the *Percentage*. Each employee's weekly daywork earnings were, then, increased by this percentage (Roethlisberger and Dickson, 1939/1975: 409). This entails that if all the hourly wages of the employees increased the percentage would just be smaller, just as it entails that if only a few employees' daily wage increased, the percentage to everyone would be smaller. To support this incentive system, a production norm was generated by the management as something to which the employees could aim at achieving (Roethlisberger and Dickson, 1939/1975: 410). The incentive, however, was defeated by the employees' introduction of a group norm on a *fair day's work* (Roethlisberger and Dickson, 1939/1975: 415; Mayo, 1949/2000:70). The output reference norm was collectively sanctioned by the employees in the sense that employees who exceeded this collectively agreed norm was frowned upon (Roethlisberger and Dickson, 1939/1975: 420). This sanctioning of those who exceeded the output norm manifested itself psychologically as teasing and the creation of nicknames, and physically as *Binging* – a game in which one employee would hit another hard on the upper arm. The problematic interrelation between norms on input and norms on outcome is clearly stated in the dialogue between two employees (worker-8 [W<sub>8</sub>] and worker-6 [W<sub>6</sub>]):

W<sub>8</sub>: (To W<sub>6</sub>) *“Why don't you quit work? Let's see this is your thirty fifth row today. What are you going to do with them all?”*

W<sub>6</sub>: *What do you care? It's to your advantage if I work, isn't it?*

W<sub>8</sub>: *Year, but the way you're working you'll get stuck with them.*

W<sub>6</sub>: *Don't worry about that. I'll take care of it. You're getting paid by the sets I turn out. That's all you should worry about.*

---

<sup>9</sup> The hourly rate was determined as a weekly average of the employee's output (Roethlisberger and Dickson, 1939/1975: 410).

*W<sub>8</sub>: If you don't quit work I'll bring you.*

(Roethlisberger and Dickson, 1939/1975: 422)

The employee in this example who exceeds the output norm is not considered by the second employee as someone who is doing something which might bring about a collective benefit in form of an increase of the daywork rate, because he is violating the output norm. Like in the example from Lysgaard's (1961/2010) study, the employee who exceeds the output norm is considered as someone who intentionally desires to differentiate himself. The interesting here, nevertheless, is that employee who exceeds the output norm has nothing to gain besides the pride or intrinsic motivation he might obtain from his work and a small adjustment of his daywork rate, because his extra effort is not positively appreciated by his peers. The employee who is exceeding the output norm is, rather, viewed as egoistic insofar as he is not considering the effects his actions might have on the other employees in terms of layoffs, cutting of rates, or higher production standards (Roethlisberger and Dickson, 1939/1975: 418).

As these two examples demonstrate, the individual who differentiate herself may as a consequence be psychologically and physically punished by her group members. Just as the agent may be ostracized from the group, and as a consequence suffer a loss of belongingness. This has also been experimentally confirmed in recent studies on social punishments, which demonstrates that participants often punish high achievers (Cinyabuguma, Page, and Putterman, 2006: 273), just as they tend to punish other social agents who deviates positively from a group norm, because this *atypical* behavior is viewed as antisocial (Irwin and Horne, 2013: 568). In that sense, the employee who chooses not to pursue an opportunity may simply refrain from doing so, because the costs associated with doing so exceeds the monetary outcome she might receive from doing it. This might, then, be viewed as the clash of two systems of norms, one belonging to the employee-organization relation and another belonging to the employee-colleague relation. Insofar as what is considered to be *right* in one relation cannot necessarily be considered as right in another, because to do what is right in one system, the employee had to violate the rules of another. In that sense, from a vertical perspective the outcome may be fair, however, from a horizontal perspective the actions which

brought about the event may be viewed as if the employee violated a social norm on equal input. The question may then be can an outcome which is indirectly brought about by an action which is considered wrong be fair? In other words, the interaction between cognitive structures governing fair pay differences and those governing intra-group pro-social behavior may; on the one hand, undermine the perceived fairness of the pay system; just as it, and perhaps more importantly, on the other hand, might undermine social relations. I return to this in chapters VIII and X. What is proposed here, is in other words, a different perspective on the causality of the reward process – a perspective which emphasizes the indirect causality of differentiation as analytically distinct from the direct causal relation between managerial decisions on rewarding some behavior and the actual act of rewarding, insofar as the former is related with horizontal relation, just as it temporally precedes the other (the act which is rewarded necessarily precedes the practice of rewarding, although it might, of course, be extrinsically motivated), while the latter is solely related with the vertical relation.

## **2.5 Tentative Conclusions**

Total compensation systems based on the two values, individuality and individualization, may be characterized by a focus on recognizing individuality by allowing the employees to design their own compensation packages to their satisfaction, and by personalizing the employment relation by paying the individual and not the job through a greater focus on the individual employee's skills, knowledge, and competencies; and by a focus on recognizing the individual's contribution through, on the one hand, person-based base pay, and, on the other hand, through a focus on performance-related-pay.

A total compensation system based on these characteristics may, then, lead to a particular conception of horizontal employee-colleague relations, insofar as these might be characterized by inter-group differentiation due to the individualization of the employment relation; and intra-group total pay differences due to the focus on paying the individual and performance-related-pay. The *primary or formal drivers* of intra-group inequality may, in other words, be recognition of individuality, individual capabilities, and performance; while the *secondary drivers* of intra-group inequality might be differentiation through how the individual composes her compensation package; differentia-

tion through the career opportunities she pursues with regard to personal development opportunities; and differentiation through performance. The primary drivers are associated with the organization's compensation policy, and as such exogenous to the employee-colleague relation, while the secondary drivers are associated with the individual behavior caused by these drivers.

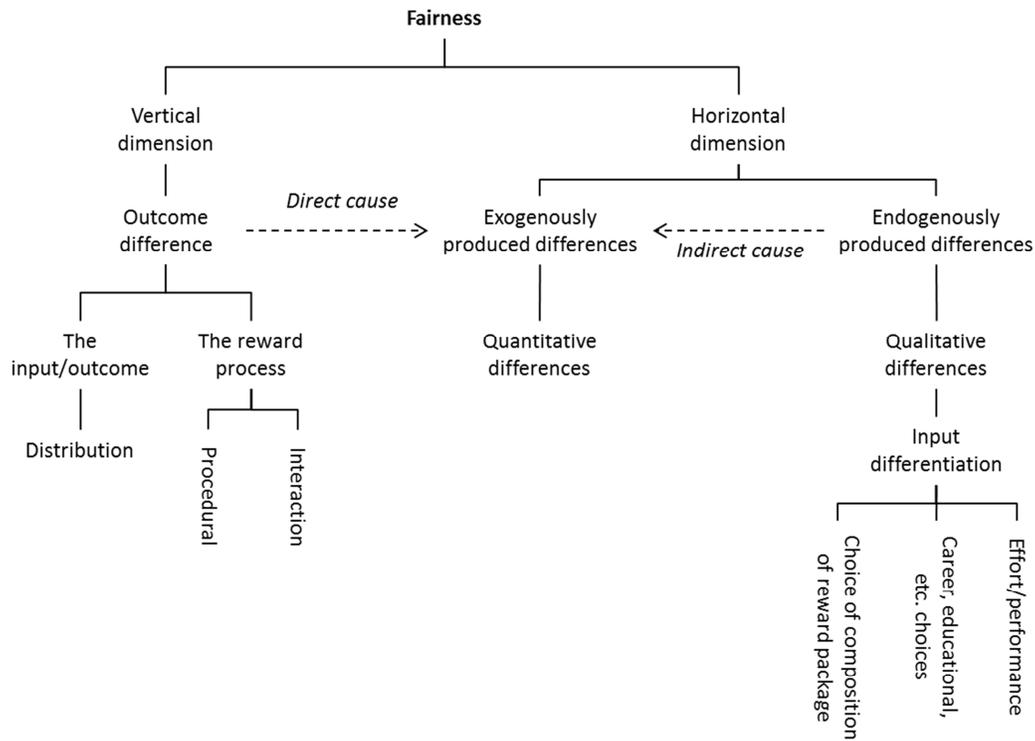
Whether or not the activated behavior is recognized is a matter of vertical fairness in the employment relation. Fairness concerns related with equity can be met with policies on openness, insofar as it allows the employee to rationally assess the distributional fairness, while procedural fairness concerns can be met with policies on inclusion, openness about the foundation of the assessment, and the right to raise objections – that is, policies on voice and inclusion. This leads to the assumption that the rational employee is able to assess the fairness of her own outcome and her colleagues' outcome if provided with the necessary information. Moreover, this rational assessment also leads to rational acceptance of outcome differences. However, is this acceptance, solely, based on *cold* rational assessment of technical fairness? Is the rational acceptance called for here not also influenced by social emotions and moral concerns? That is, can fairness in horizontal relations be reduced to a matter of technical concerns, and if not does this entail that fairness is irrational?

From a different perspective, one might also inquire into the social nature of these secondary drivers of intra-group inequality insofar as these shifts the cause of inequality from exogenous to endogenous, to the horizontal employee-colleague relation. That is, inequality is not something *done to the group*; rather it is caused by the behavior of the group members, the individual employee must therefore be willing to differentiate herself. Differentiation may, as demonstrated above, come with a price, insofar as the better performer may be punished or ostracized from the group – both social phenomena may lead to a loss of belongingness.

From what has been argued here, it seems that solidarity cannot be reduced to matters of quantitative outcomes, insofar as endogenous qualitative differences are introduced with the personalization of the employment relation which also seems to influence the sense of a common fate. In addition to this, it also seems to follow that horizontal fairness cannot be reduced to matters related with outcome, but must also take

into consideration input, insofar as differentiation is also related with input, while difference is related with outcome. Difference, however, causally follows from acts of differentiation if individuality and individualization is taken seriously as foundational values of compensatory practice. Fairness, then, from a horizontal perspective cannot be reduced to a matter of solely exogenously introduced differences, insofar as these are brought about, indirectly, by endogenously produced acts of differentiation – that is, fairness is not just related with the results of reward processes and, as such, reducible to the vertical relation. A tentative model on the relations here described can now be produced:

Figure 3: Hierarchical Taxonomy of the fairness relations described in the chapter.



The basic argument pursued here is, then, that when evaluating the fairness of a compensation system, one cannot solely rely on matters related with the vertical dimension, just as one cannot solely take into consideration matters related with either the outcome in itself or processual considerations. By taking into consideration the horizontal dimension, one may also take into consideration fairness issues related with input – matters related with intra-group norms on endogenously produced acts of differentiation.

## Chapter III

### Rationality, Irrationality, Fairness, and Individuality

#### 3.1 Introduction

In the previous chapter I discussed how contemporary theories on compensation systems may be viewed as being centered upon the two values, individuality and individualization. Here it was also discussed how changes in compensation management theories may lead to two conceptions of inequality – qualitative inequality related with input (choices regarding different opportunities open to the employees) and quantitative inequality related with outcome – implying that a theory of fairness must not only take into account these two kinds of organizational inequality, but also their causal origin, insofar as outcome inequality is causally introduced exogenously to the horizontal relation by the organization's reward practice, whereas qualitative inequality is causally introduced endogenously by the group-members themselves when they choose to differentiate themselves, pursue different goals, etc. As I also argued in the previous chapter, it is only the former mentioned causal relation which can be viewed as direct, insofar as the employees may pursue different strategies of differentiation, but these need not be, necessarily that is, transformed into quantitative outcome differences because this is contingent upon them being recognized by the organization. Furthermore, rationality was discussed as important in regard to the formation of outcome expectations and in regard to fairness as rational acceptance based on an assessment of the technical foundation of outcome decisions.

#### 3.2 Introduction to Fairness Theories

From the conception of fairness developed in the previous chapter, fairness may have two relational bases – vertical and horizontal. As I will argue later, both bases can also be analyzed from a causal perspective by the introduction of a counterfactual model. In accordance with the philosophical perspective developed in chapter I, if fairness is causally analyzed, one needs to understand the causal mode of production underlying the production of fair acts, just as one needs to take into consideration the horizontal

and vertical structures of the object of interest. Horizontally, one may distinguish between *distributive fairness* (concerning resource distribution) associated with, for example, equity theory; *procedural fairness* (concerning the decision-making process) associated with, among others, the work of Gerald Leventhal (1980: 35) and his perspective on a non-legal approach to procedural fairness; and *interactional fairness* (concerning interpersonal treatment) associated with, among others, the work of Robert Bies and Debra Shapiro (1987: 201) and their discussion of causal accounts related with the enactment of the procedures (Colquitt, Greenberg, and Zapata-Phelan, 2005: 5). These distinctions may even be related with different historical periods of research, insofar as the fairness research in the 1940s to the mid-1970s primarily concerned distributive fairness; the fairness research in the late 1970s to the mid-1990s primarily concerned procedural fairness; and the fairness research in the late 1990 to now, primarily concerns interactional fairness, and an integrative approach advocating an overall approach to fairness (Colquitt, Greenberg, and Zapata-Phelan, 2005: 7; Ambrose and Schminke, 2009).

The scientific field of organizational fairness can also be understood as centered upon a distinction between: *instrumental* models of fairness advocating a consequentialistic perspective on fairness, implying that judgments on fairness are based on a calculation of the future instrumental consequences of some action – that is, social agents want to have control over their outcome, and if this is impossible they want to ensure that they, at least, have influence on the process or the decision-maker (Thibaut and Walker, 1978: 550; Greenberg, 1990: 403; Shapiro, 1993); and *relational* models advocating a socioemotional perspective in which fairness is a matter of psychological factors such as respect, dignity, esteem, etc., implying that an action is judged as fair, if it indicate the social group's or some authority's respect – that is, an unfair treatment takes on a symbolic meaning of being rejected or excluded from the group with which the social agent identifies herself (Tyler and Lind, 1992; Tyler and Blader, 2003; Cropanzano and Ambrose, 2001: 124; Cropanzano, Byrne, Bobocel, and Rupp, 2001: 173, 174). While the first model focus on the material dimension of gain and losses, the second model focus on gains and losses in terms of esteem, thus one might argue that the first model focus on the material dimension of the distribution of tangible outcomes, while the second focus on the distribution of intangible psychological outcomes. Com-

mon to both models, according to Folger and Skarlicki (2008: 32) and Cropanzano, Goldman, and Folger (2003: 1019-1020), is that the agents described are rational in nature implying that: (1) the agents are instrumentally rational; (2) the agents perform conscious cost/benefit analysis as foundation of their choices – engage in egoistic optimization; and (3) following from (2), fairness is related with the treatment of the individual agent herself and not the well-being of others – that is, the theoretically constructed social agents do not care about the treatment of others or if they do it is solely because it has some influence on how they are treated themselves.

This critique can be viewed in accordance with the *moral turn* in organizational fairness associated with the authorship of Robert Folger (1998: 16) who argues that fairness theory have lost its moral foundation. According to Folger (1998: 13, 14), a third kind of distinction may also be called upon, insofar as one can distinguish between egoistically and morally founded theories of fairness – equity theory, for example, as will be discussed at some length later, can be categorized as a theory on distributive fairness, based on a calculative (instrumental) understanding of fairness in which the individual will mentally *keep score* and assess the relation between her input and outcome, just as this model is primary centered upon individual egoistical fairness (is it fair for me?). Centered upon this idea of mental score keeping, fairness becomes, primarily, a matter of *asocial preferences* held by the individual, thus by invoking a conception of fairness as a matter of different relative relations between input and outcome (ratios), one easily comes to underestimate the social nature of fairness – that is, the fact that social life unfolds itself in the presence of other social agents and that states of fairness and unfairness are the results of social interactions (Folger, 1998: 14). That is to say, social agents care about what and how much they obtained (distributive instrumental concerns), just as they care about why they obtained what they did in terms of relational motives (procedural/interactional relational concerns). Whether framed as material outcome concerns or socioemotional concerns about esteem, the base of fairness is self-centeredness, just as it seems to have lost its ethical and moral base (Folger, 1998: 17).

Based on a similar line of reasoning, Cropanzano, Byrne, Bobocel, and Rupp (2001: 175) and Cropanzano, Goldman, and Folger (2003: 1019) argued that besides the instrumental and relational distinction, a third variant must be taken into consideration,

the moral perspective associated with, for example, Folger's (2012) work on *deonance*. According to Folger (2012: 125), deonance is derived from the Greek word *deon* meaning duty or that which is binding. With the concept of deonance, Folger (2012: 126) addresses the duties and obligations of social life, that is, the normative *ought* which accompanies certain events. Deonance can, in other words, be understood as the conscious/unconscious restriction or reduction of autonomy which follows from acting in accordance with morality, and as such it can be related with the *not-free* actions, those which one cannot choose not to follow, the considerations which one needs to take into account, etc. – it describes, thus, the constraints which are perceived as a necessary part of social life, the constraints which are necessarily accepted, and as such it describes a distinct way of social life, one that is limited by other social agents' expectations and one's private principles (Folger, 2012: 127, 128; Folger, Ganegoda, Rice, Taylor, and Wo, 2013: 907). Deonance is a process theory, thus it focuses on the motivational state of deonance – the motivational force of moral norms – and as such, deonance captures the restrictions of one's freedom which are perceived as necessary and expected, thus in that sense, it captures a state of willingly restriction (Folger, Ganegoda, Rice, Taylor, and Wo, 2013: 908). In that sense, deonance describes a particular psychological state associated with the arousal in certain situations of certain beliefs about relevant moral rules (Folger, 2012: 126). As I will argue below, Folger's conception of deonance has its intellectual roots in a Kantian approach to morality, although the rational dimension of Kantian morality associated with practical reasoning seems to be less emphasized. Folger (2012) is also inspired by more recent work on deontological moral philosophy associated with, among others, the works of Sieghard Beller (2008; 2010) and Patricia Greenspan (2007).

According to Beller (2008: 306, 309; 2010: 123), deontic rules refer to the social rules which define the rights, obligations, and duties associated with taking part in some particular social event – what is socially accepted. Deontic rules, then, are the normative constraints of action, aiming at directing or guiding the actions of social agents in a particular direction which is in accordance with some superordinate interests – for example, in the interests of society in general. Deontic reasoning is reflection upon what is prohibited (may-not) and what is permissible/allowed (may), and on what is mandatory/required/necessary (must), what is not (must-not), and what is indifferent – allowed

but not required (contingent) – with respect to a set of social rules (Beller, 2008: 306; 2010: 123, 124). These rules along with the individuals' ability to reflect upon them are important means to understand human social cognition – that is, how individuals' make sense of the social world facing them, their relations with other individuals, and the rights and obligations inherent to these. From this conception of deontic rules it also seems to follow that permissible actions possess qualities which are positive in themselves, and as such, these positive qualities are in themselves *good* reasons to perform the action. Greenspan (2007: 175) argues on this point that while this kind of reasoning assigns positive qualities to set of actions – the permissible actions – along with negative qualities to a different set – the banned actions. In accordance with this, the agent does not, necessarily, have to be motivated to perform the actions due to the positive reasons, insofar as she might just not want bring about the negative consequences associated with the banned set of actions. In a non-trivial sense, what Greenspan (2007: 179, 181) is here arguing, seems to be that while some actions are related with purely positive qualities, and as such, do not provide any reasons against not performing them, just as they do not state negative qualities associated with performing a different set of actions – one's reasons for not performing the rational actions – those within reason – cannot be based on first-order reasons (reasons directly related with the action), but must be based on second-order reasons (reasons related with principles held). That is, by introducing a second-order level reasons, enables Greenspan to argue that agents have within their set rights, the right not to perform action – in a sense then, what she is arguing is that they hold the second-order right to give themselves permission to say *no* – and this no, which is the result of a second-order set of rights held by the agent, comes to represent a decision based on second-order reasons. At this second-order level of reason, then, reason is intrapersonal in the sense that the criticism directed at the first-order reasons (the discounting of first-order qualities) originates from the standpoint of the agent self and not some other agent (Greenspan, 2007: 185).

Turning back to Folger (2012: 127), deonance can be contrasted with *reactance theory* which described the opposite state – reluctance to accept limits of one's freedom or the unwillingness to accept the influence of other individuals as a restriction on one's freedom of choice understood as the possibility of doing what one might want to do (Folger, Ganegoda, Rice, Taylor, and Wo, 2013: 908). To explicate the difference be-

tween the two psychological states, one might argue that deonance is the motivational force invoked when bounds of morality are unauthorized sought loosened, whereas reactance is the motivational force invoked when one's freedom is unauthorized sought tied (Folger, 2012: 128). Consider, for example, the difference between the two following statements: "Don't do that!" and "You shouldn't do that, because...". While the first statement may be viewed as a limitation of one's freedom, the second one may not, because it is linked with a justification (expressed in what follows *because*), at least not insofar as this justification is perceived as legitimate (Folger, 2012: 129). Folger (2012: 129) goes on, based on this distinction between deonance and reactance, arguing that within deonance theory, legitimacy is related with deontic constraints associated with behavior and justification with consequentialistic constraints associated with outcome – that is, deonance theory focus on the interaction between behavioral constraints and outcome constraints. What Folger is trying to capture here is, perhaps, the fact that some outcomes may be justifiable within a given context, but the actions needed to bring them about may not be legitimate, just as some actions may be legitimate but their outcomes not justifiable.<sup>10</sup>

Viewed in relation with the particular conception of social life, Folger (1998: 26; 2012: 127) is approaching a particular conception of autonomy – one that is bounded – extended between “what I’m allowed to” and “what I’m not allowed to”; “what I owe you” (my obligation) and “what you owe me” (my right/claim); authority (my right to influence) and accountability (the responsibility accompanying this right to influence); etc. If this is viewed in relation to the conception of fairness given in the previous chapter, one might argue that bringing about particular outcome states (a state of difference and a state of differentiation) is justifiable with reference to, for example, the organization's compensation policy and the formal obligations inherent to the employment relation, whereas the legitimacy of the actions needed to bring these states about are governed by norms inherent to the horizontal relation. That is, because the employee cannot be directly blamed for bringing about a certain outcome, she might be blamed for performing a particular action regardless of whether or not this actually led to the present state, I return to this in chapter VI. In some sense then, this model is based on the idea

---

<sup>10</sup> This way of reasoning is common to the business ethics field and probably has its roots in Milton Friedman's (1970) essay “The Social Responsibility of Business is to Increase its Profits”.

that social agents can and will act morally, in the sense that they will act fair because it is *simply* the right thing to do. From this, it also seems to follow that fair and unfair states are something brought about – that is, they can be viewed as something which emerges within a social context and unfolds within social interaction, and as such implying not just a causal explanation of why something occurred, but also a moral explanation – a moral justification which not only come to guide how one should respond, but also a moral foundation of future actions (Cropanzano, Byrne, Bobocel, and Rupp, 2001: 165; see also chapter VI).

Simultaneously with the theories on organizational fairness, in a cross-over perspective between experimental economics and experimental social psychology, Ernst Fehr and his colleagues discusses and analyzes fairness from a motivational and social perspective (see chapter IV). The basic argument here seems to be that the normative foundation of vertical relations cannot be reduced to economic self-interest, because social agents also care about being treated fairly, again a fair treatment cannot be reduced solely to matters of outcome, but must also include a non-instrumental dimension, here identified as, for example, reciprocity. From an instrumental perspective, the judgment on the fair treatment may also be based on a distinction between envy and guilt – a distinction between “what is fair for me” and “what I owe some other – what is for her” inherent to Fehr and Klaus Schmidt’s (1999) theory on *inequity aversion* (see chapter V). In this perspective, fair behavior can be viewed as deviations from the standard model on economic rationality and selfishness, insofar as the social agents hold social preferences. This distinction between fairness and economic rationality can be viewed as emerging, in the economic approach at least, from the writings of Daniel Kahneman and his colleagues who in the mid-1980 developed a cognitive normative fairness model. The basic argument here is that the cognitive normative structuring of what is considered as fair deviates from economic rationality, insofar as it is based on a cognitive normative foundation emphasizing, for example, hedonic adaption and temporal relations between past and present, rather than economic rationality (see chapter VII). While the perspective developed by Fehr and his colleagues take on a *hot* approach based on emotions, social and moral norms, Kahneman and his colleagues take on a *cold* cognitive approach. I return to these authors in later chapters, hence it suffices as a short introduction.

The phenomenon of fairness in the employment relation is, then, studied in both economics and organization theory. Common to both fields seems to be that fairness is viewed in opposition to self-centered rationality, although the assumption on rationality is still present in the relational and instrumental approaches to organizational fairness. That is, while social agents care about being treated fairly, their motives are egoistic in nature and associated with either gaining control of outcome or due to socioemotional reasons. In opposition to this, Folger and his colleagues proposed a moral turn in organizational fairness based on the argument that fairness need not be reducible to rational motivation nor should it be separated from its moral foundation. In the cross-over between economics and cognitive psychology, Kahneman and his colleagues demonstrate that the cognitive normative rules on which judgments on fairness are based differs from the rules governing behavior in economic rationality. Fehr and his colleagues, later, demonstrate that the social normative foundation of fairness may differ from the assumptions in economics. From this it seems to follow that one might distinguish between: models based on pure economic rationality without fairness consideration and models which take into consideration fairness; the fairness dimension may also be dissolved into a distinction between: fairness models based on egoistic rationality and fairness models based on a normative foundation different from egoistic rationality; and lastly, the fairness models based on a normative foundation different from egoistic rationality may be dissolved into a distinction between: fairness models based on, for example, morality, fairness models based on a distinct cognitive normative logic, and fairness models based on social/moral normative logic. The scientific field may, then, be dissolved into distinctions associated with motivation.

However, besides these distinctions on motivation, one may also distinguish between models based on rational judgment and models based on bounded rationality or limited cognitive abilities – fairness is also related with an epistemic dimension stemming from epistemic problems associated with producing true (cognitive rationality) and right (evaluative rationality) judgments – that is, epistemic problems related with the very foundation of judgments in terms of, for example, the collection of information (for example, experience) and the cognitive processing of this. Kahneman's conception of fairness also implies limited cognitive abilities, and more recently, *Fairness Heuristic Theory* focusing on the formation of judgments from a relational perspective has been

developed. Judgments on fairness, according to this perspective, can be viewed as related with the collection of information, the assessment of the symbolic meanings of this in terms of a socioemotional sense of inclusion/exclusion from the group, and, subsequently, these symbolic inferences drawn from the information come to determine how the social agent addresses the group and interprets later actions done to her by that group or authority (for an overview of this theory see, for example, van den Bos 2001). Fairness heuristic theory, then, is centered on the idea that social agents do not formulate novel judgments on every event which might call for it, because that would imply that the social agent should stop and reason about all the potential implications and concerns at all these events, just as the information this would require is seldom available to the agent (Lind, 2001: 60; Cropanzano, Byrne, Bobocel, and Rupp, 2001: 170). Hence, one might also distinguish between fairness theories which assume rational judgments, and theories which assume limited cognitive abilities.

From a metatheoretical perspective, then, theories on fairness seems to have emerged in opposition to rationality, and as such it might be argued that in order to fully comprehend fairness theories one needs to understand the rational perspective – in a negative sense as what fairness is not or at least in opposition to Folger's (1998;2012) theory on deonance. In contrast to Greenspan's (2007) perspective on deontology, Folger (1998; 2012) reduces the influence of rationality. However, the aim here is not, as such, to reconstruct the theoretical field of fairness theories, rather the aim is to understand its genealogical foundation in rational philosophy. Besides this, one might also argue that the critique of rationality provided by these theories is too narrow, insofar as rationality cannot be reduced to economic rationality or instrumental rationality. Moreover, rationality cannot be separated from theories on morality, insofar as morality generates the social dimension of rationality. The aim, here, is not, as such, to vindicate rationality; rather the aim is solely to argue that fairness cannot and neither should it be separated from its rational philosophical genealogical foundation. In addition to this, it might also be argued that if the theories discussed above on vertical fairness has their point of departure in rational philosophy, a perspective which has as its aim to understand horizontal social fairness as an extension of the vertical perspective, must also have its point of departure in rational philosophy. In that sense, a second aim of this chapter is also to understand the moral foundation of horizontal social relations – that is,

in juxtaposition to the vertical perspective, the horizontal perspective is not related with asymmetric power relations, insofar as the employee-colleague relation is assumed to be based on status equality. In a similar vein, it must also be taken into consideration that in opposition to the vertical perspective, the employees have no actual responsibility to each other besides that inherent to the relation itself. Thus, the aim of this chapter is to analyze and discuss the relationship between morality and rationality, just as the aim is to understand the basic moral foundation between social agents. The approach here is diachronically structured in a manner which focuses on the theoretical evolution of the relation between rationality and morality, and synchronically structured in a manner which allows one to understand how the classic perspectives are related to some relevant modern perspectives.

After having explicated the moral foundation of horizontal symmetric relations, the perspective, in chapter IV, shifts to a focus on the on the formal and informal foundation of vertical relations in order to understand the normative foundation of these. Based on this understanding of the normative foundation of horizontal and vertical relations, the perspective, in chapter V, shifts to a focus on relational fairness centered upon a distributive fairness perspective in order to understand how the general claims/rights derived in chapter III may be influenced by claims related with relational differences stemming from input differences, just emotions such as envy or guilt might also influence the nature of the claim. Having analyzed and discussed the normative nature of vertical and horizontal relationally derived claims, general as well as outcome related, the perspective, in chapter VI, shifts to a focus on the cognitive foundation of causality of blame in order to understand the difference between direct and indirect causality. Centered upon this discussion and analysis of the normative relational foundation combined with the cognitive foundation of causality, the perspective, in chapter VII, shifts to a focus on the relation between temporality and hedonism, in order to add the influence of the temporal structuring to the already identified normative and causal framework. So far, the focus have been on understanding the structuring of difference, in chapter VIII, the perspective shifts to a focus on differentiation and the social norms and psychological needs which might influence the functionality of this.

### 3.2 On Rationality and Morality

In the following section, I will address the relation between morality and rationality. To secure the underlying chronology of the diachronic approach, the point of departure will be Plato's philosophy. Following this, I will move to the work of Thomas Hobbes and his conceptualization of instrumental rationality, psychological egoism, and his conception of passions within reasons. Besides this, Hobbes also argued that the relation between the individuals is centered upon a social contract, thus the contract perspective will be introduced and its relation with James Coleman's modern conception of social relations. From this discussion on the social contract and based on the philosophy of David Hume, an early conception of social norms as the foundation of social life will be introduced. From the utilitarian perspectives of Hobbes and Hume, the analysis moves on to a discussion on Adam Smith's philosophy as a middle point between utilitarianism and Kantian philosophy. Immanuel Kant's philosophy can be read not just as a critique of utilitarianism, but also as a continuation of Platonic philosophy. After this, I introduce the philosophical perspective associated with Jeremy Bentham's approach to utilitarianism. From the classic authors, the text moves on to discuss two modern approaches to rationality, Jon Elster's and Raymond Boudon's respective perspectives.

#### 3.2.1 Plato

In book four of his work *Republic* (380BC), Plato claims that the human soul, what one might call the mind or *psyche*, consists of three masters, appetite (in Greek *alogiston* meaning non-calculating), reason (in Greek *logismos* meaning reckoning, calculation, or computation, and in Latin *rationem* meaning understanding, motive, or cause; and *ratus* meaning to reckon or think), and the high spirit (in Greek *thymos* meaning spiritedness).

Plato (380BC/1961a) argues that the appetite (derived from Latin *appetitus* meaning desire toward and *appetere* meaning to long for or desire) is the root of the basic desires (food, water, sex, etc.), and he adds that these desires are qualitatively neutral because they do not refer to, for example, a *good* drink, rather they refer to universal opposites, for example, thirst refers to the desire for a drink (object) – to drink (action). One might agree with Plato that when one is thirsty one does not long for a good drink, but solely a drink. However, one may, on the contrary, argue that when one desires something it is because one considers it to be *good* (Lesses, 1987: 150). A similar, but

also somewhat different, interpretation is proposed by Christine Korsgaard (2008: 104; 2009: 137), who argues that this qualitative neutrality follows as a logical consequence from his understanding of the human psyche as consisting of three masters, because if the appetite could rule on its own as a non-cognitive appetite of already qualified objects it would imply that these objects were essentially qualified by something besides the rational calculations of the agent, thus leaving no space for reason. Hence, when it was argued that the agent must consider it to be good, this should be understood as being qualified as such by the agent herself – or more in line with Plato’s understanding, as being qualified as such by the rational dimension of the appetite on behalf of the entire soul (Kahn, 1987:85). In other words, one should distinguish the non-cognitive hedonic experience of the object from its *true* nature as *good* or *bad* in terms of evaluative rationality, albeit in terms of cognitive beliefs it might be added that one should classify the object as drinkable (Kahn, 1987: 85). In addition, the appetitive part also holds beliefs on how best to obtain its end – practical reasoning (Kahn, 1987: 86) – for example money (or acquisitiveness) is pursued because it is causally linked to the obtainment of other desires (Lesses, 1985: 151; Plato, 380BC/1961a: 580e).

Not all pleasures, of course, are good nor are all pains bad, and sometimes these becomes mixed as in the case of *Schadenfreude* when the individual experiences the mental pleasurable pain of *gild* mixed with joy (Plato, 360BC/1961b: 46c). Michel De Montaigne (1580/2003) captures this ambiguity, when he discusses the complexity of human nature:

“Both in public and in private we are build full of imperfection. But there is nothing useless in Nature – not even uselessness. Nothing has got into this universe of ours which does not occupy its appropriate place. Our being is cemented together by qualities which are diseased. Ambition, jealousy, envy, vengeance, superstition and despair lodge in us with such natural right of possession that we recognize the likeness of them even in the animals too – not excluding so unnatural a vice as cruelty; for in the midst of compassion we feel deep down some bitter-sweet pricking of malicious pleasure at seeing other suffer.” (Montaigne, 1580/2003: 892)

Plato considers this ambiguity of the senses in his dialogue on emotions and evaluative judgment, “*Philibus*” (360BC/1961b). His basic argument can be reconstructed as fol-

lows; it is impossible to deduce a value judgment from the non-cognitive experience (for example from emotions) of something – that is, it does not follow, logically at least, that just because something feels *good* it is also *good*, for example in a moral sense (Plato, 360BC/1961b: 13a-b). Hence, he argues that the metaphysical foundation of justice cannot be the hedonic experience of good or bad. Additionally, it is important to recognize the connection that Plato constructs between emotions and hedonic experience; because emotions, here, are not solely considered as non-cognitive qualitative representations of the connection between the internal condition of the agent and the external world facing her, but also as being accompanied by hedonic experience in themselves (Plato, 360BC/1961b: 39a), as it was the case, in the previous example on the pain of Schadenfreude.<sup>11</sup>

Plato argues in the book 10 of Republic that this conflict of emotions emerges because individuals are split between an impulse to follow their strong emotions, and a wish to follow the path of reason and law (Plato, 380BC/1961a: 604a-d). The rational part in the human is measured in feelings (Plato, 380BC/1961a: 604a), and it is calm and temperate (Plato, 380BC/1961a: 604e). The irrational urge to flee into emotions make humans “...*stumbling like children, clapping one’s hand to the stricken spot and wasting the time in wailing...*” (Plato, 380BC/1961a: 604c). Plato goes further and argues that it is “...*the part of us that leads us to dwell in memory on our suffering and impels us to lamentation, and cannot get enough of that sort of thing...*” (Plato, 380BC/1961a: 604d).

Plato criticizes poetry for its tendency to invoke these emotions, that is, its tendency to nurture this inferior part of the soul. What he criticizes again is that these emotions distort the understanding and as such hinders the individuals from pursuing true values – values reached through reason. These emotions idles the mind, they act as comfort and as shields against the brutal facts of the real world. These emotions when they are strongest make individuals follow impulses; they make the individuals pursue

---

<sup>11</sup> This particular understanding of emotions is also found in Aristotle (350BC/2007: 1378a; 113) and Spinoza (1677/1996: II/153; 79) who both focus on the hedonic dimension of the emotions themselves. For some reason, modern phenomenological philosophers and sociologists such as Martin Heidegger (1926/2007: 165), Jean-Paul Sartre (1939/2002: 39), and Niklas Luhmann (1994/2000: 316) seems to favor the first dimension of emotions as a social phenomenon which emerges as an internal reaction on something external.

something which appears to be good, just as they make them refrain from pursuing other objects out of fear. Montaigne (1580/2003) makes similar argument:

“For in his case the impress of the emotions does not remain on the surface but penetrates through to the seat of his reason, infecting and corrupting it, he judges by his emotions and acts in conformity with them.” (Montaigne, 1580/2003:49)

Emotions then are false impressions, like the ones generated by the senses, and the ones generated by hedonic experience. As such like the emotions described by poets this emotional experience is nothing but imitations of what is good and bad, they are not true values. Montaigne (1580/2003) argues in a similar manner that emotions cannot be viewed as true perceptions:

“For while it is true most of our actions are but masks and cosmetic, and it is sometimes true that:

Behind the mask, the tears of an heir are laughter.

Nevertheless we ought to consider when judging such events how our souls are often shaken by conflicting emotions. Even as there is said to be a variety of humours in our bodies, the dominant one being that which normally prevails according to our complexion, so too in our souls: although diverse emotions may shake them, there is one which must remain in possession of the field; nevertheless its victory is not so complete but that the weaker ones do not sometimes regain lost ground because of the pliancy and mutability of our soul and make a brief sally in their turn.”(Montaigne, 1580/2003: 263)

Emotional experience might sometimes hide the true value of an object. For example, a particular object may appear more desirable if for some reason it hold some sentimental value, just like those individuals who choose to dwell on their past or their hope for what the future will bring them (Montaigne, 1580/2003: 11). Emotional experience like hedonic experience, make individuals follow impulses. Additionally, it also changes the values associated with particular objects to generate satisfaction. Plato, however, argues in similarity with his idea that individuals should choose true beliefs that:

“And shall we not say that the part of us that leads us to dwell in memory on our suffering and impels us to lamentation, and cannot get enough of that sort of thing, is the irrational and idle part of us, the associate of cowardice?” (Plato, 380BC/1961a: 604d)

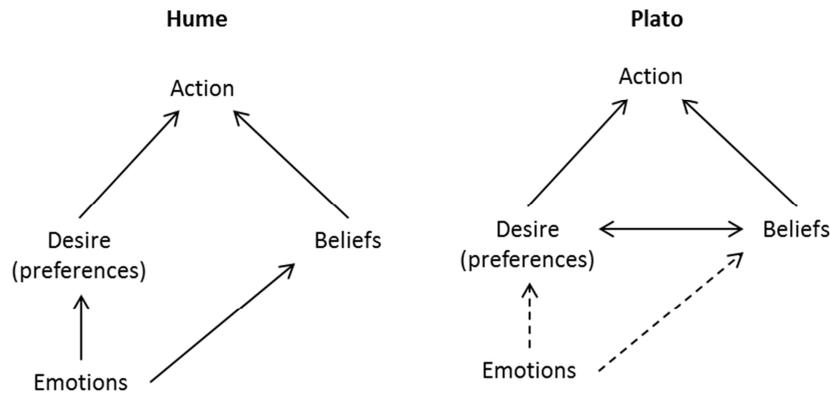
According to this view, individuals who dwell on the past or are prone to succumb to follow their emotional experience are cowards, like those who choose to hold false beliefs. They do not, as rational actors, act on the factual world. However, it may be argued that Plato in his use of reason as opposite to emotional experience generates a false proposition. Because, as Hume (1739/1985) argues, can emotional experience be true or false, does not this logically follow from an existence of something on which this can be measured? Emotional experience, like hedonic experience, as such is not reflections of something which exists in any objective manner; they are original and bear as such no representative quality which should render them true or false. Of course, one might argue that it opposes reason if it distorts the practical beliefs or if it generates false beliefs. Hume (1739/1985) argues:

“First, when a passion, such as hope or fear, grief or joy, despair or security, is founded on the supposition or the existence of objects, which really do not exist. Secondly, When in exerting any passion in action, we chuse means insufficient for the designed end, and deceive ourselves in our judgment of causes and effects. Where a passion is neither founded on false suppositions, nor chuses means insufficient for the end, the understanding can neither justify nor condemn it.” (Hume, 1739/1985: 463)

“A passion is an original existence, or, if you will, modification of existence, and contains not any representative quality, which renders it a copy of any other existence or modification. When I am angry, I am actually possest with the passion, and in that emotion have no more a reference to any other object, than when I am thirsty, or sick, or more than five foot high.”(Hume, 1739/1985: 463)

The distinction between Plato’s and Hume’s conceptions may be illustrated as follows:

Figure 4: Causal diagram of rationality (inspired by the illustrations made by Elster, 2009b: 15)



In the Platonic approach (on the right hand side) the rational action is caused in the right way when the action is produced by a reasonable desire (the connection between desire and beliefs), while irrational desires are caused by emotional influence (the connection between emotions and desires). In the Humean approach (on the left hand side), desires are not connected with beliefs (or if they are, he would call them preferences not desires) – only with emotions, hence a desire caused by the emotional connection cannot be opposite the reasonable desire since this does not exist.

What is discussed here, is different intentional states, in the sense that the agent can intend to *do* something (want), just as she might *believe*, *desire*, *hope*, or *fear* something. The intentional verbs, in italics, refer to different intentional states; the agent may, for example, want to do something – to bring some particular state about or some value into action as a particular value expressed, albeit symbolically, in her course of actions, just as she might, from an epistemological perspective, believe that it is snowing outside, desire it to show outside, hope that it snows outside, or fear that it snows outside. These different intentional states, in other words, expresses different psychological modes related with the same propositional content – it is snowing outside. Consider, for example, the statement: “Sandra believes Mont Blanc to be the highest mountain in Europe”. The proposition of belief is the content and not the attitude, insofar as the attitude refers to the object of the belief, Mont Blanc. That is, the attitude expressed in the psychological modes refers not to the proposition, but to the object, and as such, the fact that Sandra believes something cannot be expressed as in the relation: Sandra-*believes*-Mont Blanc is the highest mountain in Europe, or in general terms: subject-

intentional mode-proposition; rather the relation is: subject-intentional mode-object. In that sense, in the example, Sandra is not expressing an intention to believe, but solely a belief, and as such, the intentional does not refer to the act of *wanting to believe* (below I will argue that this is a self-defeating act insofar as one cannot make oneself believe at will), but the act of *wanting* Mont Blanc to be the highest mountain in Europe – the intentional mode refers to the object. Based on this, one might argue that a belief is true insofar as it expresses a correct representation of the object, the aim here, then, is to secure an *acceptable* relation between mind and world. Whereas epistemological statements can be tested drawing on a binary distinction between true and false, psychological modes cannot, inasmuch as the causal pattern is opposite to the one characterizing true/false statements – in the sense that when one is expressing a belief about the world, in an epistemological sense, one is following a mind-to-world causal pattern because one is trying to create a relation between mind and world, whereas if one hopes something, one is following a world-to-mind causal pattern because one is trying to *change* the world – make it into something representing one’s hopes or desires.

Thus, Plato might be right when he argues that intentional states influence the subject’s apperception of the world, in the sense that the intentional mode refers to characters of the object. Just as Hume might be right when he argues that one should not test psychological modes based on a true/false binary schematic, insofar as these do not express anything about the world. Based on a similar logic, if a right action should exist it would presume the existence of a general form/standard, something which the individual could attune her action to – on the other hand, if the action is taken solely to express an intentional state, the aim of the action is to bring some state about, to attune the world to one’s intentions, and as such, not attune one’s action to some general form. As I will argue below, Plato did not propose a model of action based on external validation, rather he proposed, like Kant, a model based on internal validation based upon attuning one’s actions with the unity of one’s soul, these desires which the action should be attuned to should, however, be in accordance with reason – that is, based on true experience of the world and one’s own desires, just as the unity of the soul is produced through reason. In that sense, it is this act of internal validation that Hume is opposing, because it presupposes that reason can be cause of action. From a Platonic perspective, one could also argue that to be rational is to choose to live in the real world, to be cou-

rageous enough to choose the real world above some not world of pretend, and to be courageous enough to do what is right – to pursue a right action and to act in accordance with it.

Besides the basic desires, Plato also claims that motivational irrationality stems from the high spirit, defined as “... *that with which we feel anger...*” (Plato, 380BC/1961a: 439e) without it being calculated by reason (Plato, 380BC/1961a: 441c). The high spirit refers to the higher level passions, the desire for fame, honor, respect, justice, love, beauty etc. As such the high spirit must be understood as the values, standards, ideals or principles held by the agent. Again, as stated in relation to the basic desires, these higher level desires also influences the evaluation of the qualities of some object, however, these evaluations are not building on hedonic experience, but rather on kind of emotional experience, and as it was the case with basic desires these also produce practical beliefs on how best to obtain them. Plato argues on the relation between the basic and higher level desires:

“And they, imitating him, received from him the immortal principle of the soul, and around this they proceeded to fashion a mortal body, and made it to be the vehicle of the soul, and constructed within the body a soul of another nature which was mortal, subject to terrible and irresistible affections – first of all, pleasure, the incitement to evil; then, pain, which deters from good; also rashness and fear, two foolish counselors, anger hard to be appeased, and hope easily led astray – these they mingled with irrational sense and with all-daring love according to necessary laws, and so framed man.” (Plato, 360BC/1961c: 69c-d)

As this quotation states, agents follow ends produced by the basic and the higher level desires, they follow these ends because they are considered to be good by the individual part of the mind. In some cases agents act on these desires because their rational part misidentifies, or perhaps is overpowered, what is good (Kahn, 1987: 88). As argued by Montaigne in quotation above, these desires are imperfections of human nature, because they make the individuals act in non-rational ways – in ways not guided by calm deliberate reason. Plato would not object to this, however, he believed, contrary to Montaigne, that individuals could gain control over these urges through reason – not make them disappear – because what makes a desire irrational is not the desire as such, but

that it may be based on a *wrong* evaluation of qualities of the object, and only through reason could the *right* qualities be revealed (Kahn, 1987: 88).<sup>12</sup>

This high spirit can be pictured as being between reason and basic hedonic desire, and as such it may ally with both.<sup>13</sup> Plato argues on the relation between the high spirit and reason:

“Would not these two, then, best keep guard against enemies from without, also in behalf of the entire soul and body, the one taking counsel, the other giving battle, attending upon the ruler, and by its courage executing the ruler’s design?” (Plato, 380BC/1961a: 442b)

“...his high spirit preserves in the midst of pains and pleasures the rule handed down by the reason as to what is or is not to be feared.” (Plato, 380BC/1961a: 442c)

That is, desires may guide the agent towards solely pursuing her hedonic pleasures momentarily activated in her mind or it may help her to overcome an immediate temptation. For example, the appetitive part may argue; “what is the harm in taking another drink”, while reason argues “you had enough – it is not good for you”, and if the individual holds the value “responsible drinking” she may be able to get past this temporary preference reversal. This, simple, example thus illuminates the motivational conflicts which emerges as a kind of a motivational paradox – when the agent pursuing some end she evaluates as both good, in terms of the hedonic pleasure of the present moment, and bad in terms of reason and value-commitment, albeit these two may have different ends in mind. For example, the agent above may have held the value “no honor in drinking” and the result would have been the same. What Plato is arguing, it seems, is that individuals can act contrary to reason and contrary to their values; which might lead to an emotional conflict. However, Plato did not conceive this high spirit as a means of self-punishment, as a Freudian interpretation might suggest in regard to his idea of the Super-Ego (Kahn, 1987: 83), rather he saw it as an additional element guiding the actions

---

<sup>12</sup> It could be argued that Plato considers a distinction based on reason – scientific logic – between *right* and *wrong* judgments, whereas Montaigne used a distinction based values between *good* and *bad* or perhaps *vices* and *virtues*.

<sup>13</sup> This interpretation of the relations between reason, appetite, and thymos is based on various quotations cited in this section, and as such I do not take into consideration the differences in the assumed psychological foundation which may exist in the various works of Plato. Korsgaard (2008: 104), for example, solely considers Plato’s definition of thymos in his “Republic”, therefore she argues that it always allies with reason, as it is also stated in the quotation above.

of the agent and her evaluative judgments as outward projected in relation to the object of her desire and sometimes in competition with the other desires. Plato argues on this matter:

“And do we not, said I, on many other occasions observe when his desire constrains a man contrary to his reason that he reviles himself and is angry with that within which masters him, and that as it were in a faction of two parties the high spirit of such a man becomes the ally of his reason?” (Plato, 380BC/1961a: 440b)

For example, an agent who smokes may know the dangers stemming from her habit, but still has no desire to stop – reason alone generates no motivational element, a point which I will return to below in relation to Hume’s conception of rationality. However, if for some reason the agent develops a set of values stating that she wants to live a healthy life, this may generate the missing motivational element. In other words, there is a difference between *knowing* and *wanting*. This is also captured in the figure above by the double connection between beliefs and desires, and the connections: desire-action and belief-action.

So far, in this text, I have considered the irrational parts of mind as claimed so by Plato. To Plato, reason was also a desire, a desire to learn and for wisdom (Kahn, 1987: 81). Furthermore, only the rational part of the mind has the capacity to take into account all the relevant and available beliefs and information – that is, to pass judgments based on *all things considered* foundation. The irrational parts, on the contrary, only take into the account the beliefs about how best to obtain their particular ends (Lesses, 1985: 155). This double role which Plato assigns to reason as; on the one hand, a desire and part of the psyche; and on the other hand, as the agents capacity to unite her different desires (illustrated in the figure above by the two-sided connection between desire and belief) – I return to this double sided role below. To understand Plato’s distinction between the rational and the irrational parts of the psyche consider the following quotations:

“And is it not the fact that that which inhibits such actions arises when it arises from the calculations of reason, but the impulses which draw and drag come through affections and diseases?” (Plato, 380BC/1961a: 439d)

“...we claim that they are two and different from one another, naming that in the soul whereby it reckons and reasons the rational, and that with which

it loves, hungers, thirsts, and feels the flutter of titillation of other desires, the irrational and appetitive..." (Plato, 380BC/1961a: 439d)

"Does it not belong to the rational part to rule, being wise and exercising forethought in behalf of the entire soul, and to the principle of high spirit to subject to this and its ally?" (Plato, 380BC/1961a: 441e)

"But wise by that small part that ruled in him and handed down these commands, by its possession in turn within it of knowledge of what is beneficial for each and for the whole, the community composed of the three." (Plato, 380BC/1961a: 442c)

As it is argued in the first quotation above, the irrational parts of soul are those which arise through impulses, as actions which motivation stems from external events, which, as such, are not authored by reason. The rational part, Plato adds in the second quotation above, is that which reckons and reasons, and as such it is *giving* reasons, albeit not through impulses, but through the rational part of the psyche, it is the *slow* and thoroughly calculation (i.e. reflections on advantages and disadvantages) and deduction of reasons from the epistemological foundation of *all things considered*. The irrational parts of the mind are rapidly changing, adapting to present environment and the present desires – an understanding which can be illustrated by Montaigne's metaphorical representation of human action as a chameleon:

"Our normal fashion is to follow the inclination of our appetite, left and right, up and down, as the wind of occasion bears us along. What we want is only in our thought for the instant that we want it: we are like that creature which takes on the color of wherever you put it. What we decided just now we will change very soon, and soon afterwards we come back to where we were: it is all motion and inconsistency." (Montaigne, 1580/2003: 374)

In the following two quotations, Plato focus on the second role of rational reason, that is, rational reason as the reason of reasons, as a kind of second order reason which act as the author and the legislator of the constitution of the mental unity, and as such it acts on behalf of the entire soul by *orchestrating* the principles of desires (below I refer to these as; first, second, and third order desires) around a central principle of unity (i.e. harmony). Korsgaard (2008: 101; 103; 105; 2009: 134; 137) refers to this as the *Constitutional Model* of rational action, and what characterizes this conception of rational action, she argues, is that reason is viewed as the author of the constitution (and as the constitution itself) which unites the soul, therefore when the agent follows the principle of reason it is not because reason won its battle with the appetite, but because the prin-

ciple was in congruence with the constitution of the soul. Hence, deliberate actions (rational actions) are recognized as such, because they sprung from the constitution of the soul, and as such they are representations of the unity of the soul, and of the rational individual – that is, they are not just fragments representing parts of the agent’s identity, but her whole or united identity as a rational agent.<sup>14</sup> In that sense, the rational agent not only chooses her actions, she also chooses her ends – at least insofar as she chooses between which ends to pursue. Korsgaard (2008: 101) contrast this model of rational action with what she calls the *Combat Model* of motivational conflict – a conception of the human psyche which describes the different desires as battling against each other, within the mind of the agent, and the desire which *wins* is that which rules, although this may be how it is sometimes imagined, it seems, at least, problematic because it would require agents without essence or identity, and as such without any foundation of rationality. I return to this below in the section on Humean rationality.

After this clarification of the double role rational reason, I will now turn to the practical function of this second-order level rational reason in relation to practical reason. Plato argues on reason as the foundation of choice:

“I mean in choosing, greater lengths and avoiding smaller, where would lie the salvation? In the art of measurement or in the impressions made by appearances? Haven’t we seen that the appearance leads us astray and throws us into confusion so that in our actions and our choices between great and small we are constantly accepting and rejecting the same things, whereas the metric art would have canceled the effect of the impression, and by revealing the true state of affairs would have caused the soul to live in peace and quiet and abide in the truth, thus saving our life?” (Plato, 380BC/1961b: 356d-e)

Plato argues in the quotation above that if reason is made the ruler of choice it will through reason induce consistency in the individual’s preference because it enables a conscious choice between greater and smaller, that is, it enables differentiation between the two, thus rendering it possible to choose the one the individual may desire. It is this very act of differentiation between objects possessing more or lesser value which makes the choice in itself rational, because it is not based on appearance, either emotional experience, hedonic experience or sensory perception, it is based on abstract and complex

---

<sup>14</sup> This particular model of rational action is also represented in Immanuel Kant’s approach to rationality, which I will discuss below.

reflection. Plato goes further and argues that rationality also tells the individual which to choose:

“Again, what if our welfare lay in the choice of odd and even numbers, in knowing when the greater number must rightly be chosen and when the less, whether each sort in relation to itself or one in relation to the other, and whether they were near or distant? What would assure us a good life then? Surely knowledge, and specifically a science of measurement, since the required skill lies in the estimation of excess and defect – or to be more precise, arithmetic, since it deals with odd and even numbers.” (Plato, 380BC/1961b: 356e-357a)

The right choice as opposed to some inferior choice should be distinguished using rational calculation or knowledge based on evidence instead of experience. To understand this distinction and its reference to mathematical (or logical) reasoning it is important to recognize the etymology of rationality: ratio-nality is derived from Latin *rationalis* meaning belonging to reason, and *ratio* meaning to reckon, calculate or reason – in short to apply logical reasoning. Applying mathematics as a metaphor for rational reason seems intuitive, because mathematics is a closed logical system based on certain procedures, rules and properties. Every number has certain properties which must be taken into account, just as the limited number of mechanical procedures based on certain rules limits the space of possible actions. That is, when solving an equation one has to respect the properties of the parts, just as one has to follow a certain procedure to get to the right result, hence it also limits the space of possible outcomes. Another quality of mathematics is that it produces consistent results due to consistency in methodology which implies that an infinite number of individuals can solve the same equation and reach the same result because they follow the same rules. What this seems to imply is that rationality, for Plato at least, was the study of practical logic, that is, a study of how humans reached the result they did in judgment as well as in action. Because mathematics is a pure logical system it may have inspired Plato, as a rationalist, to search for the same logical structure in human conduct, that is, finding the *right* properties of the input (evaluative rationality) and finding the *right* procedures to reach the *right* result (practical rationality). The study of the right properties of objects should not be based on experience in any sense, because this varies from individual to individual and from situation to situation, it should rather be based on pure universal forms, because not only did these properties signify the real qualities of the particular object, they were also univer-

sal in nature, which implies that they were constant throughout time, and between individuals. This way reasoning is often referred to as mathematical Platonism (Balaguer, 1998) and is, of course, more complex than how it was described in this section, because it is related to existence of universals<sup>15</sup> and its consequences for ontology and epistemology – this however, is beyond the scope of this discussion.

To understand the process of choice, in terms of mathematical Platonism, it is necessary to design some rather fixed qualities that most individuals would use in their process choice, hence Plato choose the dichotomies; good and bad, right and wrong etc. These should, ideally, be based on universal forms freed from personal experience. This is a kind of decision rule which made it possible for Plato to continue his search, just like rational choice theory today is based on certain decision rules, I return to this below in relation to the underlying preferences of rational choice.

As I showed above, in matters of practical rationality reason might be a better judge; however Plato did not consider reason alone or aided by high spirit as the only motivational elements. Because as is argued in the quotation below, the rational part can also produce counter beliefs of causal connections and thereby turning the pleasurable desire into something painful.

“And knowing that this lower principle in man would not comprehend reason, and even in attaining to some degree of perception would never naturally care for rational notions, but that it would be especially led by phantoms and visions night and day - planning to make this very weakness serve a purpose, God combined with it the liver and placed it within the house of lower nature, contriving that it should be solid and smooth, and bright and sweet, and should also have a bitter quality in order **that the power of thought, which proceeds from the mind might be reflected as in a mirror which receives likeness of objects and gives back images of them to the sight, and so might strike terror into the desire when, making use of the bitter part of the liver,** to which it is akin, it comes **threatening and invading,** and diffusing this bitter element swiftly through the whole liver **produces colors like bile, and contracting every part contorting the lobe and closing and shutting up the vessels and gates causes pain and loathing.**” (Plato, 360BC/1961c: 71a-b; my emphasis)

---

<sup>15</sup> The interested reader can consult Plato’s original philosophy of the logic of forms see, for example, the dialogues: *Cratylus* (here Plato discusses the problem of knowing the forms); *Symposium* (here Plato discusses the particular form of beauty); *Sophist* (here Plato discusses the essence of a form); and *Philebus* (here Plato discusses the unity problem). For a thorough, albeit challenging, discussion of mathematical Platonism, its consequences and relations to modern analytical philosophy, see Mark Balaguer’s (1998) book on the subject.

What Plato is discussing here is the basic physiological desire; this quotation, however, demonstrates the relation between the rational part and the desires. The content of the rational judgment is prescriptive when it threatens and commands, and evaluative because it is concerned with what is good and bad, beneficial and harmful. However, it does not command using solely the means of reasons; rather it induces qualities in the object which make it appear less desirable. In other words, the rationally induced desires are of objects whose qualities or value are imputed by the rational part of the mind, and projected through the mind, and when reflected upon the object, it generates within the agent a feeling of pleasure or pain, desire or fear. This statement should be viewed in relation to the second-order role of reason I discussed above, because it presupposes the existence of the idea of what is *truly* desirable, and as such it requires the agent. Even though, this quotation might seem a bit peculiar, it should not be problematic to imagine that agents create these symbolic connections, and as such imputes a symbolic value onto the object.

What is also implied, it seems, is a distinction between first and second order desires. A first-order desire can be defined as a mere want or as a proposal made by the appetite (Korsgaard, 2009: 141). A second-order desire, on the other hand, can be defined as the answer of reason to the proposal of the appetite – whether to act or not – or in a less metaphorical language, as a desire not to desire what is desired (Elster, 2000: 21). In other words, the agent may want something, but not wanting to want it, and if this conflict is to be resolved, something is needed to bind the agent from these things she want, but do not want to want. Again, the relation between the three parts of the soul needs to be taken into consideration, just as the double role of reason must be called upon to solve the motivational conflict.

As the Platonic perspective seems to imply, one cannot comprehend rationality as derived from reason or desire in isolation. The rational action is caused by reasonable desires and instrumental beliefs. Reason adds a positive dimension to desires, insofar as these must be truly desirable, that is, these must be based true impressions guided by reason based inquiries into their foundation. Irrationality can, then, be associated with actions based on desires not governed by reason; rather these desires may be motivated by emotional experience. However, rationality is not the dispassionate cold process of

deliberation insofar as reason in isolation only provides advice on certain matters and secures balance between the different parts of the soul. I return to the relation between emotions and irrationality in section 3.3.7.

### **3.3.1.2 Social Rationality**

Plato, of course, did not solely consider rationality in regard to material and external objects, but also in regard to internal motives, and its relation to the foundation of morality. Social rationality can be defined as the rationality of social actions, or equivalently, as the social rationality of actions<sup>16</sup>, although the latter definition needs a social foundation which, as such, was not present in the philosophy of Plato. Social rationality as a concept, then, covers the social actions of social agents, and the social foundation of their actions, and as such it is closely related to morality and ethics, however, it can, thus, be viewed as an extension of the rational morality associated with Plato (and Kant), even though it also takes into account the social context. Consider Plato's description of rational morality and the role of reason:

“But the truth of the matter was, as it seems, that justice is indeed something of this kind, yet not in regard to doing one's business externally, but with regard to that which is within and in the true sense concerns one's self, and the things of one's self. It means that a man must not suffer the principles in his soul to do each the work of some other and interfere and meddle with one another, but that he should dispose well of what in the true sense of the word is properly his own, and having first attained to self-mastery and beautiful order within himself, and having harmonized these three principles, the notes or intervals of three terms quite literally the lowest, the highest, and the mean, and all others there may be between them, and having linked and bound all three together and made of himself a unit, one man instead of many, self-controlled and in unison, he should then and then only turn to practice if he find aught to do either in the getting of wealth or the tending of the body or it may be in political action or private business – in all such doings believing and naming the just and honorable action to be that which preserves and helps to produce this condition of the soul, and wisdom the science that presides over such conduct, and believing and naming the unjust actions to be that whichever tends to overthrow this spiritual constitution, and brutish ignorance to be the opinion that in turn presides over this.”  
(Plato, 380BC/1961a: 443c-e; 444a)

---

<sup>16</sup> Social rationality as a concept and theory, probably originates from George Herbert Mead's (1934/1962) discussion of a social equivalent of the Kantian categorical imperative. Mead's perspective, however, is, as such, not in complete coherence with Kant's perception, and this might be the reason why this essay is very seldom referred to.

Plato's moral psychology is built upon his conception of the second-order role of reason as uniting the desires of the three parts of soul under one constitution. The aim of this second-order reason is to identify the best way of generating a harmonious living, a life in which the individual becomes who she is, and stays true to her own way. Rationality in general, and social rationality in particular, are both the foundation of deliberative action, and when the agent acts, her actions are representations of this rationality and the uniting constitution of the three parts of the soul. As such, the process of deliberation unites the three parts of the soul in its reflexive process of deciding what to do by orchestrating the proposal of the appetite with the commands of reason, and its values – thus, this orchestration of the three parts of the soul as a united system also constitutes the identity of the agent (Korsgaard, 2008: 119; 2009: 179). An agent acting entirely on her own principles might be called authentic, because she lives a life guided by her own reasons, desires, beliefs, and values. Such an agent has integrity, because she stays true to her own principles, and she applies these principles to resolve her inner motivational conflicts. That is, agents may sometimes do something not because they want to but because they want to want to, as sometimes seen in acts of character planning, or because the agent believes others would want her to do it. What this implies, then, is that while reason may provide the individuals with second-order desires it also requires a third-order principle of desire, namely, that the individual actually want-to-want-what-she-want. Following this, it might be argued that such an individual actually is acting in her own self-interest. If by reason all the false images have been removed, then rationality steps in to find the best possible way of unifying the three parts of the soul. This also shows that morality is not the opposite of rationality, that is, rationality understood as pursuing one's self-interest. Because, according to the principle of third-order desires, the agent who acts in a just way must do so because she wants to – that is, she has chosen to be just, and the action she has identified as just is one which unite her three parts of the soul, and as such, it represents the constitution, just as it symbolizes the constitution of the agent as just.<sup>17</sup> This principle, then, rejects the contractual view of justice stating that agents engage in these contracts not because they want to be just, but rather because they fear themselves being wronged (Plato, 380BC/1961a: 359a). According to this principle, it is rational sometimes not to act in one's own interest if the result of this

---

<sup>17</sup> This particular of understanding of social rationality and morality is also represented in the Kant's perspective which I will discuss below.

is better than the one which could have been obtained by acting solely in one's self-interest.<sup>18</sup> Moreover, this also states that individuals are purely driven by their desires, and that the only purpose of rationality is an instrumental one. To do what seems to be the best thing to do is, then, not a priori rational it becomes so if and only if the individual actually want to do it, that is, it does not disturb the inner harmony of the soul – it is in tune with nature of the individual, that is, her third-order desires.

Another important quality following from this harmonious state of being must be that the agent has adapted her aspirations not only to achieve internal, but also external consistency. Because, the agent torn between conflicting desires must be in a similar state to one who wants something but cannot get it. The first case is a matter of the agent's power of will, whereas the second one is a matter of the will of others. In the first case, the agent may act in accordance with her principles, even though the will has limits to its power or she may design personal norms. However, according to Plato this would mean that she is acting in a manner which is not in tune with her nature, that is, she is not rational, at least not in regard to the Constitutional Model of rational action. Moreover, the agent who desires something which is out of her reach may also be torn between her desire and her possibility of obtaining it, to resolve this conflict the individual may change her preferences to obtain cognitive resonance – that is, design new third-order preferences which are more appropriate to her situation. However, then it might be argued that she is no longer acting in accordance with her true being, or perhaps this is what Plato accounted for when he argued that the individual should get to know herself. This lack of self-knowledge may then be a source of irrationality?

In accordance with what has been discussed so far, it seems that three different kinds of rationality can be distinguished in the philosophy of Plato (Kahn, 1987: 86-87). The weak kind of rationality is *purely* instrumental, and the role of reason is solely to identify the best possible way of obtaining its exogenous goals, and to distinguish between right and wrong practical beliefs. The semi-strong kind was presented above; reason is practical, it aims at finding the best possible combination of the different desires stemming from the three parts of the mind, and based on some idea of the good life, a harmonious life. The strong kind of rationality is when rationality rules, when

---

<sup>18</sup> This may be viewed as an early account of the Prisoners' dilemma.

reason is not only practical but also theoretical and engaged in identifying the pure forms.

From what has been argued it also seems to follow that if Plato is correct, rationality cannot be reduced to self-interested actions, at least not in its conventional meaning as stating that the agent only performs actions which springs from her private self-interest, and as such only conforms to social conventions if it is within her interest to do so – a perspective which is, however, inherent to Thomas Hobbes' conception. Rather, the rational agent conforms to social conventions because she wants-to-want, in the sense that she tries to secure a balance between extrinsic stimuli and intrinsic motivation. That is, like the Kantian agent, I will discuss below, the Platonic agent is not *forced* to social rationality, rather she commits herself to these demands by acting in a way which is reasonable – governed by reason, and balanced between her desires and the desires others. The model described here, in some sense at least, can be viewed as some of the, and in combination with Kant's perspective the entire, philosophical basis of Folger's (1998/2012) theory on deonance discussed above.

A first take on the relation between fairness and rationality would then be to argue that fairness cannot be reduced to matters of reason alone, because the rational agent will also have to commit herself to the values implied in such a compensation policy – meaning that she must view these values as being in accordance with her morality, just as she, of course, must view the decision as being based on true conceptions, for example, adequate measures. Solely believing in the technical adequacy does not lead to rational acceptance – only if accompanied by a belief in the values implicitly/explicitly stated values can acceptance be viewed as rational – as a belief in the rightness and truth of the policy. Furthermore, it might be argued that Plato's conceptualization of the relation between rationality and morality is *asocial*, insofar as his perspective, as such, is not related with how the individual should act towards other individuals. Morality is, here, not directed towards other individuals, but is solely related with how the individual should govern herself, how she obtains a moral way of life – a balanced life. Although Kant, as I will argue later, proposes a similar model, he does include a reference to other individuals as, at least, some of the foundation of morality.

### 3.3.2 Thomas Hobbes

Even though Hume's arguments resembles the ones of Hobbes there are some differences which need to be illuminated to understand the origin of instrumental rationality, but also the nuances of Hobbes' theory on rationality. Hobbes follows Plato's line of argument in his book on the organization of the state, "Leviathan" (1651). Hobbes, like Plato, also argues that society should be organized in the image of man (Hobbes, 1651/1985: 81). Hobbes do not, contrary to Hume, believe that reason was purely instrumental (Hobbes, 1651/1985: 115), just as two different kinds of rationality can be identified in his work, a weak kind and a strong kind.

According to Hobbes, the role of instrumental reason is to identify links between actions and consequences, and the link between consequences (Hobbes, 1651/1985: 110). Identifying these causal linkages is the work of science, it is a continuous process of reasoning and as such it has not as its goal to establish certainty, but rather to pursue and update what is already known (Hobbes, 1651/1985: 113). Science differs from sense data, experience of facts, and from practical experience because:

“...Sense and Memory are but knowledge of fact, which is a thing past, and irrevocable; Science is this Knowledge of Consequences and dependence of one fact upon another: by which, out of that we can presently do, we know how to do something else when we will, or the like, another time: Because when we see how any thing comes about, upon what causes, and by what manner; when the like causes come into our power, we see how to make it produce the like effects.” (Hobbes, 1651/1985: 115)

From past experience, sense data, memory, and practical experience, the agent obtains knowledge about how things has been, what has happened etc. and using this as input, the role of practical reason is to draw inferences from the past to the future, from one consequence to another. In that sense, even though the matters of reason are practical it is based on abstract thinking. And as such it goes beyond simple learning (conditional response), and beyond learning from experience, because Hobbes also argues that it is the knowledge of what is within the power of the agent – that is, it not just about practical reasoning, the causal connection between an action and a goal, but also about recognizing the possibility or potential.<sup>19</sup> Later in his text, Hobbes defines power as the abil-

---

<sup>19</sup> There is always the possibility of misinterpretation when reading these classic philosophers because the language has evolved, however, the word “power”, in French “pouvoir” or Latin “potentia” both translations means “possibility” or “potential” and could be interpreted, in accordance with Aristotlian physics,

ity to obtain some future good, and distinguish between natural power (faculties of mind and body) and instrumental power (power as a means to gain more) (Hobbes, 1651/1985: 150). It seems, then, that practical reason also relates to the agent's space of possible actions, what she can do and the consequences they produces. Hobbes writes:

“For the Thoughts, are to the Desires, as Scouts, and Spies, to range abroad, and find the way to the things Desired...” (Hobbes, 1651/1985: 139)

The intellectual link between Lawler's (2003) conception of rationality as the foundation of action and Hobbes' original statement seems rather clear. The sense of control is important to both theories insofar as the agent must believe that it is within her power to bring a particular state about. Lawler (2003), however, distinguishes between effort-performance reasoning related with the agent's conception of self-efficacy, and instrumental reasoning related with the performance-reward connection (see the previous chapter). The latter connection, then, adds that the agent must not only perceive the action as being within her power, but also being possible to bring about given the fact that the agent cannot control her environment.

To Hobbes, like Plato and Hume, humans act because they have reasons to do so and these reasons are internal. The rational act is not just based on internal reason, it is directed towards some end, and as such it is caused by either appetite (pleasure) or aversion (pain) (Hobbes, 1651/1985: 119). Hobbes distinguishes between basic pleasures and pleasures learned from experience. Aversion, on the other hand, is not necessarily natural or experienced (Hobbes, 1651/1985: 120). Hobbes, like Plato, based his philosophy on hedonism, but argued against stable preferences in a manner like that of Montaigne:

“And because the constitution of a man's Body, is in continuall mutation, it is impossible that all the same things should alwayes cause in him the same Appetite and Aversion: much lesse can all men consent, in the Desire of almost any one and the same Object.” (Hobbes, 1651/1985: 127)

In addition to not having constant preferences, Hobbes also argues against the existence of universal goods. Here it seems that Hobbes' perspective differs from Plato's in the

---

as the knowledge of “what could be” as opposed to that “which is” (actus), and that “which is impossible”, or solid.

sense that Plato believed in the existence of a universal good, albeit this was often corrupted by false experience (see the above section).

According to Hobbes, rationality is not only characterized by a teleological pursuit of one's goals (means-end deliberation), but also by deliberation, that is, the act of carefully planning one's actions and considering their consequences (Hobbes, 1651/1985: 127). Deliberation is the act of using *foresight* to determine the consequences of one's actions and as such it is concerned with the future, expected, consequences (Hobbes, 1651/1985: 129). It is, in a sense, related with manipulating the environment in accordance with one's ends. This process of deliberation does not stop until the action is either deemed impossible or it is carried out.

Moreover, the rational action should be a voluntary action aimed at some desired end or avoiding some undesired end and it should be based on a process of careful planning considering the consequences. So far, it seems that Hobbes has not pushed his understanding past instrumental rationality. Hobbes, however, seems to argue, in his chapter on madness, that any desire, when it becomes excessive, is irrational, contrary to reason, because it hurts the individual (Hobbes, 1651/1985: 140). Even though Hobbes argues that human beings are characterized by the state felicity, a continuous pursuit of what the individual considers as good (Hobbes, 1651/1985: 130). He also argues that when this pursuit becomes excessive, compulsory, it becomes irrational, contrary to the main principle of rationality, the pursuit of wellbeing. This stronger kind of rationality, then, moves past instrumental rationality and argues that desires can be opposite to reason, albeit not because they are based on some false experience, as Plato argued, or because they are against one's moral judgment, as argued by Montaigne, but simply because they are contrary to the main principle of rationality. To some extent, this might also be what was argued by Plato, when he argued against solely following one's appetite, higher desires or reason but that one should find a combination of these. Additionally, Hobbes also argues, later in his text, that social actions which are not aimed at obtaining peace are contrary to reason, just as actions against the social contract are because they break it and reduce human society to a state of war, and as such it is against the agents' own wellbeing (Hobbes, 1651/1985: 187). The rational agent, then, should

not prefer momentarily pleasure; rather he should aim at securing continuous gains and a long term perspective:

“Felicity is a continuall progresse of the desire, from one object to another; the attaining of the former, being still but the way to the later. The cause whereof is, That the object of man’s desire, is not to enjoy once onely, and for one instant of time; but to assure for ever, the way of his future desire. And therefore the voluntary actions, and inclinations of all men, tend, not only to the procuring, but also to assuring a contented life...” (Hobbes, 1651/1985: 161)

What Hobbes is arguing is that the rational agent should not go for the short terms wins. Here Hobbes seems to capture the essence of, at least, the utilitarian approach to, social rationality, insofar as the action should be of such a kind that it does not go against the self-preservation of the agent, that is, the action should be in accordance what is reasonable in both a private and social sense. As such, the rational agent may pursue her desires within the boundaries of reasonable pleasures, and socially accepted desires. Whereas Plato’s conception of social rationality focused on how the agent should govern her life, Hobbes seems to introduce a social level, in the sense that the agent should not just govern her way of life, but also how she interacts with other social agents.

### ***3.3.2.1 Rationality and Maximization: The Two Limits of Desire***

Besides the long term perspective introduced by the state of felicity, Hobbes also argues for a principle on maximization, that is, the rational agent should not just secure continuous pleasure but also try to maximize his gains, always seek more, and compete against his previous results. Hobbes argues:

“...the cause of this is not always that a man hopes for more intensive delight, that he has already attained to; or that he cannot contend with a moderate power: but because he cannot assure the power over means to live well, which he hath present, without acquisition of more.” (Hobbes, 1651/1985: 161)

To some extent it seems to contradict the previous quotation, however, what Hobbes seems to argue is that the rational agent needs to continue developing her abilities and possibilities and that she cannot expect these to produce similar effects in the future. In that sense, the rational agent should take an active part in cultivating and developing her chances. The rational agent should compete against herself and her previous results; she

should not just passively wait for her turn, but take matters into her own hands. The rational agent is, then, responsible for her own life, and the chances she receives. An additional point which also needs to be emphasized, at least in regard to cognitive rationality is that the agent cannot expect infinite causal reactions to repeat themselves. What is important to recognize here is that Hobbes was not, as I will argue Hume was, questioning these matters from an epistemological perspective, rather it seems that he is addressing this matter within a physical and social dimension – it is a matter of power both in terms of ability – the agent is physically capable of doing it, and he masters this particular performance; and in terms of social power. If these are not properly maintained by the agent, they will lose their impact.

The rational agent described by Hobbes is an individual who always chooses her actions in accordance with her interests and her own preservation and anticipate others to do the same (Hobbes, 1651/1985: 184). Adam Smith argues in a similar way that in exchange relations the social agents should always anticipate self-interestedness and direct themselves to this particular dimension of human nature (see the section on Adam Smith below). Even though Hobbes is often ascribed the origin of psychological egoism this might be a misinterpretation (or exaggeration) because as already mentioned Hobbes did not believe that the individuals should pursue their own end at the expense of others, rather he argues that this kind of behavior is contrary to reason. In that sense, the rational agent is motivated by self-interested actions and should pursue these ends although she should also take into account that these actions has externalities, that is, these actions has consequences which affects not only the acting agent herself but also other agents (Hobbes, 1651/1985: 184). Of course, it might be argued that this too is a self-interested act, because this egoistic pursuit is only contrary to reason, insofar as it is contrary to the preservation of the agent's own wellbeing. Nevertheless, blind egoism is against reason but not contrary to human nature as described by Hobbes in his chapter on the initial position. In this description, Hobbes argues that if man is not bound by anything, neither by responsibility for him or others, has no moral code, and is not constrained by any common laws (Hobbes, 1651/1985: 186), he is reduced to a savage beast that blindly pursues his own ends at all costs. Moreover, in this initial position, because no individual is different from one another (they are equal in ability), equality of hope is born (the hope of obtaining one's goals) and if two individuals share a com-

mon goal they cannot both obtain it and as such “... *they becomes enemies; and in the way to their end, endeavour to destroy, or subdue one another.*”(Hobbes, 1651/1985: 184). The social values which permeates the initial position is competition, diffidence, and glory (see box) and the social relations between the individuals is centered upon the constant possibility of war (Hobbes, 1651/1985: 186).

Figure 5

Value	Motivation	Practice
Competition	Gain	Violence to make oneself master
Diffidence	Safety from those who has less	Defend oneself
Glory	Reputation	Value oneself

Even though this is a description of human nature, it is a distinct and abnormal situation in which the individuals are reduced to savage beasts and as such it is a dystopic description of a society without a common power, and a description of human nature if it was reduced to only pursuing its own ends, at best guided solely by instrumental rationality.

It seems that Hobbes, here, also makes an early statement on the distinction between absolute or relative gains, albeit as a distinction between competing against one or others. As already showed, the rational individual competes against herself, that is, she compares her present result to her former (intrapersonal comparison), and as such should only care about her own results. Hobbes, nevertheless, identified interpersonal comparison as a part of the non-rational human nature motivated by the two social values, competition and glory. These particular values are against reason because they lead to a state of war and as such against the primary principle of rationality, the preservation of oneself. However, contrary to both Plato and Montaigne, the war following from these desires are not solely a war against oneself (what Hobbes called madness) it also has a social dimension – that is, it has social consequences. One might argue that Hobbes by focusing on social consequences adds an important element to the discussion on the reason and desire, because he elevates the discussion from a psychological level to a social level by not only enquiring into the personal consequences of the social atom,

but actually tries to understand the collective social consequences. Hobbes might, then, be one of the firsts (perhaps the first) to separate the two levels, individual and societal level. In other words, to follow one's desires blindly is not only madness; it is also non-social and as such against reason. Albeit one could argue that this perspective is rather *cold*, insofar as social rationality solely becomes a matter of being rationally responsible – towards oneself and others.

From the discussion above, it seems to follow that the rational (in the semi-strong form) individual only engages in intrapersonal comparison, in juxtaposition to the non-rational individual who engage in interpersonal comparison. It seems that the modern discussions on this matter follow a similar distinction when it identifies the former with rationality and the latter with non-rationality or envy. The relation between social comparison and envy has its origin in Aristotle's work on emotions, "Rhetoric" (350BC/2007). According to Aristotle (350BC/2007: 146), the structural antecedent of envy is social and geographical closeness – physical and social proximity. In terms of social closeness, envy demands equality in terms of birth, social stand, moral, and resources; and in terms of geographical closeness it demands the presence of the other – that is, terms of the spatiotemporal dimension (Hobbes, 1651/1985: 146). As such, envy is built on the counterfactual belief "it could have been me" (Elster, 1999: 170), and is invoked in situations of negative comparison, that is, when those to whom the individual compares herself have more, and because they are equals this feeling of being inferior is turned inward because the individual has only herself to blame (Hobbes, 1651/1985: 147). Benedic De Spinoza (1677/1996: 88) argues in a similar way that envy is related to the competition between equals and as such it is an inner disturbance of this sense of equality which generates cognitive dissonance in this process of interpersonal comparison, see also chapter V.

In modern theory on behavioral economics, different positions have emerged on this problem of comparison. Robert Axelrod (2006: 110) argues that the rational individual should only compete against herself, thus she should not be envious. According to Axelrod (2006:110) envy emerges because the individual perceives gains and losses as related to a zero-sum game, that is, the gains of Ego is the loss of Alter. Additionally he argues that this perception emerges because, if the individuals have no objective

measure of how well they are doing, they will start to compare their results with those around them (Axelrod, 2006: 111). This strategy, Axelrod argues, is contrary to rational choice because: on the one hand, it leads to a social-destructive competition between the players; and on the other hand, it leads to self-destructive envy.

Axelrod argues, related to games on the Prisoners' dilemma, that the rational individual should use the following as standard of comparison:

“A better standard of comparison is how well you are doing relative to how well someone else could be doing in your shoes. Given the strategy of the other player, are you doing as well as possible?” (Axelrod, 2006: 111)

What Axelrod is arguing, here, is that the rational individual should compare her results to those who are equal to her and not just in matters of focal points of comparison but those in a similar situation. As such it is not about being better than the other, but doing as well as possible given the situation. This also relates to the rationality of evaluation and in particular how counterfactual beliefs and emotions affect the evaluative dimension. Axelrod argues further that in, a standard infinite symmetric Prisoners' dilemma game with simultaneous moves, the dominating strategy is Tit-for-Tat, which implies that as long as the other player is cooperation the first player should also cooperate and when the other player deviates she should also deviate (Axelrod, 2006: 112). This Tit-for-Tat strategy, then, entails that the players will always obtain equal results until one of them deviates. This position illuminates Hobbes' distinction between intrapersonal and social comparison as well as it also explicates the rational limit of following one's desires, that is, when this becomes social-destructive and self-destructive.

From this it seems to follow that Lawler's (2003) argument on rational outcome expectations being based on horizontal social comparison has its intellectual roots in a utilitarian conception of rationality. Although one could argue that Lawler (2003) neglects the reason of why this is irrational, insofar as he seems to be arguing that it is irrational because it is wrong, the utilitarian tradition, on the other hand, emphasizes that it is wrong because it is psychologically as well as socially undermining. From this, it also seems to follow that the competition which follows from performance evaluations and symbolic distinctions associated performance appraisals and public display of these ratings might undermine the rational foundation of agency. Hence, if competition is the

foundation of performance management, and horizontal social relations, the assumption on rational agency becomes crucial to the functionality of the system, insofar as an activation of irrational emotions such as envy might lead to antisocial behavior which might not only undermine the functioning of the system, but also the horizontal social relations.

### **3.3.2.2 Rationality and the Social Contract**

In that sense, the human nature described in the initial position is non-rational in accordance to Hobbes' semi-strong version of rationality. To explain the foundation of social relations, that is, the relation between rational individuals, Hobbes applied the metaphor of a social contract. According to Hobbes the social contract is mutual (symmetric) transference of rights motivated partly by desire and partly by reason:

“The Passions that encline men to Peace, are Feare of Death; Desire of such things as are necessary to commodious living; and a hope by their industry to obtain them. And Reason suggesth convenient Articles of Peace, upon which men may be drawn to agreement. These Articles, are they, are called Lawes of Nature...” (Hobbes, 1651/1985: 188)

He argues further:

“Whensoever a man Transferreth his Right, or Renounceth it; it is either in consideration of some Right reciprocally transferred to himselfe; or for some other good he hopeth for thereby. For it is a voluntary act: and at the voluntary acts of every man, the object is some Good to himselfe.” (Hobbes, 1651/1985: 191)

Hobbes distinguishes between *the right* and *the law* of nature. In relation to the former he argues that the individual is free to pursue his own ends with the means he, by judgment and reason, finds necessary; and in relation to the latter he argues that actions not aimed at the preservation of the individual, are contrary to reason (Hobbes, 1651/1985: 189). By the transferring of rights the individual becomes constrained in his pursue of ends, just as he cannot hinder the person who he transferred it to from benefitting from it (Hobbes, 1651/1985: 191). The agent is, then, obligated to adhere to this contract as long as the other person also adheres to it (Hobbes, 1651/1985: 190). Contrary to Plato, it seems that Hobbes uses this social level as a constraint on human actions and because the social level is based on the image of the individual it is as such rational – its laws are

rational. Thus, to act against the laws of society is to act against reason and, as such, rational individuals become governed by a rational social level determining what is morally right, which desires are contrary to reason, etc.. What is interesting here is the distinction between individual level and social level rationality, society is based on a rational foundation which stems from the individuals, and the individuals governed by this rational state becomes rational. The causal connection it seems is this: from individual rationality follows social rationality and from social rationality follows individual rationality. Even though Hobbes believed individuals to be rational, he also knew how fragile this state was, thus he uses the social level to secure rationality.

More recently, the sociologist James Coleman (1990) proposed a similar conception of social life consisting of an exchange of rights to act, claims, and obligations between social agents and between social agents and social institutions (Coleman, 1990: 45). Coleman (1990) proposed this view of social life in order to integrate micro-sociological theories on social interaction with macro-sociological theories on social institutions. The agent carrying out a particular action, then, is actually acting on two beliefs (1) she has the right to perform the action; and (2) she has the right to bring about a particular event. Coleman (1993) argues on this matter:

“Right” in the usage of this paper is a noun: One holds a right, gives up or acquires a right in exchange, creates or destroys a right. The same word, “right”, is used as an adjective: “He did the right thing,” or “It is right to do that.” These two usages are related: One has a right to do the right thing. If one does not have the right, it is not right, but wrong to do it.” (Coleman, 1993a: 226)

The social world facing the agent is as such a world consisting of rights (Coleman, 1990: 50). Coleman (1990) argues on this point:

“...there is no single “objective” structure of rights of control, but a structure of rights of control subjectively held by each actor in the system, and that we may speak of a right of control over an event being held by an actor only when that right exists in the subjective structures held by each of the actors affected by exercise of that right.” (Coleman, 1990: 51)

Coleman (1990: 54, 57), here, seems to be arguing that the right to control actions are held at the micro-social level, while the right to distribute these rights are collectively held, just as he also seems to be arguing that an agent only hold this right as long as those affected by the consequences of the action recognizes the agent’s right, hence

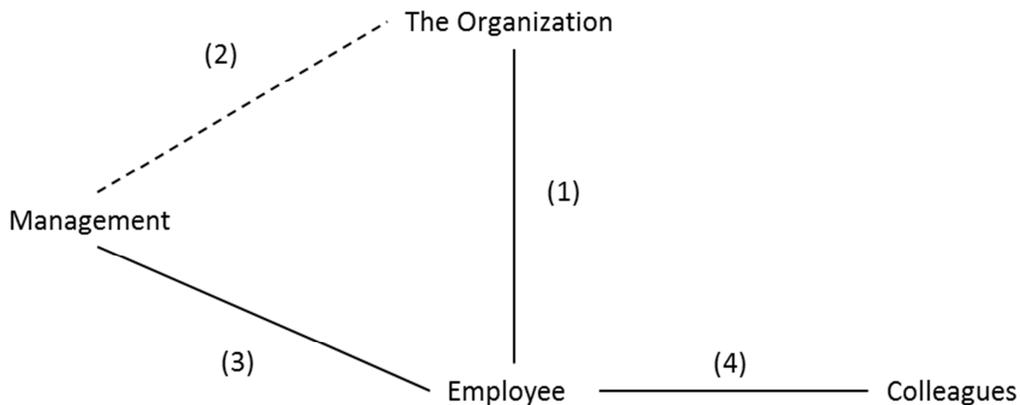
rights are only held as long as a *social consensus* exists on the particular distribution of rights. Moreover, this also implies that the social agent cannot necessarily pass a right to another social agent, because she might not have the right to do so (Coleman, 1990: 58).

In that sense, Coleman's (1990) modern formulation of the Hobbesian perspective might be interpreted as *opinion sensitiveness* – that is, by taking part in social life, the agent accepts the right of other social agents to influence her life, just as she is allowed to influence their lives. Hence, social life might be conceptualized as the exchange of rights to influence one another. By taking into consideration what others might think, the agent is limiting herself and her pursuit of private desires. The foundation of the rational acceptance of fairness is, according to this perspective, found in the interaction between the micro and macro level – that is, between social institutions and social interaction. Rational acceptance is, in other words, more than just acceptance of technical features, it is also the acceptance of one's right to claim – acceptance is, in other words, an acceptance of one's place in the hierarchy, and an acceptance of the other party's right to distribute. To enter an employment contract implies the transfer of some rights (for example, the rights to determine when and how one likes to work) in exchange of some other rights, for example, the rights to one's total compensation package. However, by entering the employment relation, the individual also enters a social contract inherent to the employee-colleague relation. The existence of such a horizontal social contract which the employee enters when enters the formal employment contract may, as such, also explain the observations from both Lysgaard's study and the Hawthorne studies discussed in the previous chapter. Hence, one might argue that the formal vertical contract governing the organization-employee relation is accompanied by a *tacit* horizontal social contract governing the employee-colleague relation. Aligning these two dimensions may, as I will discuss in the next section, support the functioning of the compensation system, while misalignment may not only undermine the functionality of the system, but also the horizontal social relations – I return to this in chapter VIII. From what has been argued here, it seems that in order to understand social life, one should take into consideration the transfer of rights, horizontally as well as vertically, which simultaneously come to govern organizational life. That is, in addition to the different dimensions of the vertical contract, which will be discussed in the next chapter,

one also needs to take into consideration the horizontal social contract governing employee-colleague relations.

Furthermore, from this conceptualization of the organization, it also seems to follow that one should distinguish between: management/employees; management/organization; organization/employees; and employee/colleagues. This can be illustrated as follows:

Figure 6: A Relational Illustration of Organizational Life.



By entering into a contract with the organization, the employee transfers a certain set of rights (1) to the organization in exchange of a certain set of claims, just as she indirectly agrees to that the organization transfer a part of these rights – the rights associated with management – to the management (2). The transfer of the right to manage generates the foundation of relation (3). Relation (4) can be viewed as a *by-product* of relation (1). Procedural fairness covers relation (1), interactional fairness covers relations (2) and (3), and distributive fairness covers relations (1) and (4), insofar as relation (4) enters into the framework through social comparison and, as such, it influences the employee’s sense of right to claim in relation (1). The perspective developed here, focuses solely on relation (4), but acknowledges the implications of the other relations.

Moreover, in accordance with Hobbes’ argument on the relation between individual and collective rationality, it might be argued, as Rescher (2002: 1) does, that out-

come fairness is contingent upon the distribution of outcome claims being fairly distributed – meaning that if a first-order distribution is in accordance with a second-order distribution of claims, then this cannot be unfair from a first-order perspective, although the principles on which the distribution itself is based might be unfair. That is, from a distributive perspective, one needs to take into consideration the second-order distribution of rights and not just the first-order distribution of outcomes. The question then becomes a matter of what constitutes a second-order claim. The distribution of second-order claim rights is also different from deservingness, insofar as it is constituted differently, hence the employee may not, necessarily, deserve the pay she is entitled to in her contract (Rescher, 2002: 2). In addition, if one adds to this framework Coleman's (1990) idea on the distribution of rights as based on social consensus, then one needs to take into consideration not just how these claims are formally constituted through contracts, job-descriptions, hierarchy, etc., but also the social constitution of these – the normative foundation. Because, as Rescher (2002) argues, the question is not whether or not the agent accepts some first-order distribution; rather the question is why social agents accept second-order distributions. Thus, the notion of acceptance developed in the section on Plato becomes important, inasmuch as acceptance implies the acceptance of the values which the distribution comes to represent – that is, the values associated with the second-order distribution of rights.

### ***3.3.2.3 The Social Contract and the Origin of the Utilitarian Perspective on Social Norms***

Hobbes argued that the principle of social rationality governing social relations could be *reduced* a rationally motivated – self-interested – action. Thus, at least to some extent, the rational agent would adhere to the social contract as long as it is within her interest to do so implying that social rationality is solely a matter of self-interest – in the egoistic and not in the Platonic sense of the word. In other words, as long as the costs associated with adhering to the contract does not outweigh the benefits. Based on a similar understanding, Coleman (1990) proposed a theory on social norms. In Coleman's (1990: 72) perspective, this transfer of rights can be viewed as a rational process in the sense that the choice to initiate the contract is based on a rational calculation of costs and benefits. If the transfer of rights is symmetrical in nature, then, the relation produces

something which is in the interest of both parties, hence no extrinsic compensation is necessary to make the parties adhere to the contractual obligations – such a relation may be called *conjoint* (Coleman, 1990: 72). If, on the other hand, the transfer of rights is asymmetrical in nature it might be called *disjoint* because here the relationship is not equally in the interest of both parties, hence extrinsic means are applied to make the parties adhere to the contractual obligations (Coleman, 1990: 72, 74). Coleman (1990) argues on this distinction:

”The terms ”conjoint” and ”disjoint” refer to the correspondence between the interests of the subordinate and the directives of the superordinate. In a conjoint authority relation the superordinate’s directives implement the subordinate’s interests. In a disjoint authority relation they do not, the subordinate’s interest must be satisfied by extrinsic means.” (Coleman, 1990: 74)

In that sense, horizontal relations between employees may be described as conjoint, whereas the vertical relations existing between management and the employee and between the organization and the employee can be described as disjoint. This distinction is also important in relation to Coleman’s (1990) conception of social norms:

“I will say that a norm concerning a specific action exists when socially defined right to control the action is held not by the actor but by others.” (Coleman, 1990: 243)

He goes on arguing on its connection with authority relations:

“By the definition of authority, this means that others have authority over the actions, authority that is not voluntarily vested in them, either unilaterally or as part of an exchange, but is created by the social consensus that placed the right in their hands.” (Coleman, 1990: 243)

What Coleman (1990) here seems to be arguing is that by adhering to a particular social norm the agent is as a consequence recognizing the right of the other party to control her actions, and unlike social contracts this recognition may not follow as a consequence of a deliberate transfer of right; rather the power of the social norm is based on the public consensus on which it is centered. In that sense, the right to control some actions is transferred from the agents to the social norm. Social norms are then purposively created to serve the interests of some beneficiary party in the sense that they benefit from the norm being observed and is harmed by its violation (Coleman, 1990: 242, 247). Consider, for example, the social norm stating “when you visit the home of another, you do not

smoke inside or at least ask for permission before lighting the cigarette” – the beneficiaries of this social views themselves as being in direct control of the action “smoking when visiting the home of another”, just as they, indirectly, perceive themselves as having the right to control the actions of the group at which the norm is directed (here smokers) (Coleman, 1990: 247). Moreover, those holding a particular norm perceive themselves as being in possession of a right to sanction other agents who violates the norm, just as they ascribe the same rights to other social agents holding the norm (Coleman, 1990: 248). This implies that if an agent holding a particular norm violates it, she might be sanctioned by either some third party, or if the agent is target of the other agent’s norm violating behavior holds the norm he might also sanction her (Coleman, 1990: 283).

The agent adhering to a particular norm, however, may not be doing so because she herself holds the norm, insofar as the choice of conforming/not-conforming becomes a matter of rational calculation in the sense that the sanction can be viewed as costs she incurs by performing a certain actions, and this cost is compared to the benefits she might obtain from either refraining from performing a possible action or from following a particular social norm (Coleman, 1990: 242, 243, 244, 286). The perspective developed by Coleman (1990), then, argues that the frequency with which certain actions are performed can be either reduced or enhanced through the implementation of material incentives which through rational calculation comes to structure the behavior of the social agents.

Before moving on to exemplify how this might lead to a crowding-in effect between incentives and social norms, some critical points needs to be addressed; (1) according to Coleman (1990) social agents solely conforms to social norms out of egoistical motivation based on extrinsic motivated self-interest, while this indeed is the case in some cases, the impartial spectator perspective discussed below also proposes that at other occasions social agents adhere to norms because they are sensitive to social disapproval a matter I return to below, just as non-material motives such as reciprocity may also be taken into account (see the next chapter). (2) Social norms are purposively created to benefit some party, while this might be the case with many norms; some norms also have adverse consequence or make all parties worse off (Elster, 2009: 196). Karl-

Dieter Opp (2001: 237) argues against this critique of Coleman's (1990) perspective when he argues that it is not against the perspective that a norm should have unfavorable consequences for some members of a group, just as this approach should not be considered as a functionalist approach, because these norms does not emerge because they have favored consequence, but solely because creating the social norm enables the group to reach its goals by implementing rules designed at making them better off. (3) The very existence of a social norm is sufficient to secure beneficial relations, again this might often be the case, but social norms also themselves becomes public goods like the one they were aimed at securing – for example, cooperation norms may secure the existence of some public good, however, the norm itself, cooperation, might become a public good, and as such norms supporting the norm must be implemented led to a kind of infinite regress; (4) According to Coleman's (1990) perspective, social norms are aimed at enhancing or reducing the frequency of certain actions due to their beneficial or harmful consequences, and as such adhering to social norms liberates the agent from moral reasoning, while this might be right at some occasions it would entail that all norms were in accordance with some existing moral code, just as it would entail that social norms existed for all possible actions and that immoral actions were only considered as such due to their instrumental consequences, Boudon's (2001) perspective seems to be less rigid in nature inasmuch as it primarily views the role of social norms as something which enters into the agent's own moral reasoning (see the section on Adam Smith); (5) in accordance with Boudon (2001) one might argue that Coleman (1990) does not, explicitly at least, address the cognitive dimension of the action – that is, the agent do not just adhere to a social norm, rather she does so because she perceive it as the right thing to do, albeit not in terms of the potential consequences the norm might have; and (6) in accordance with Boudon (2001) and Elster (1999; 2007; 2009), norm following cannot be captured within the utilitarian account of emotions in which the violation of a social norm simply is viewed as a cost incurred by the agent.

Coleman (1990: 437; 1993b: 529) viewed *Quality Circles* as an important new form of constitution within the organization because it generated, almost, ideal conditions for solving the incentive problem (agent/principal problem) due to: (1) the responsibility of production have been transferred from management to the workgroup; (2) each member of the workgroup is made responsible for the quality of the product and

they have the right to return a product if it does not fulfill the specified criteria, and stop production – in a way the authority distributed to the group mimics the production flow and enables short feedback lines compared to more traditional ways of organizing (Coleman, 1993b: 530); (3) besides autonomy and responsibility it is also based on a third right:

“...the right to extra pay (paid as a bonus) depending on quality and quantity of output by the workgroup.” (Coleman, 1993b: 529)

He continues this argument:

“By making benefits (income, autonomy, responsibility) contingent on the quality and quantity of what is produced, the organization creates a direct link between effort and reward.” (Coleman, 1993b: 531)

According to Coleman (1993b), the formal constitution of the organization with its distribution of rights and incentive system enables the creation of social norms within and between workgroups which are aligned with the organization’s goals. Coleman (1993b) argues further on this point:

“The innovation creates a *social group* in whose interest it is to encourage high effort and careful work on the part of the members. Social norms come into being to reinforce organizational goals. These norms go beyond overcoming incentive problems of hierarchical organization. They provide extra incentives of their own, because each member’s work benefits all. And this comes about because rights that in the hierarchical organization are held centrally and delegated vertically are instead given to subgroups...” (Coleman, 1993b: 531, italics in original)

What Coleman (1993b) seems to arguing, here, is that these norms which arise from the organization’s constitution come to support and is supported by the material incentive structure designed by the organization – that is, within these workgroups, social norms arises on, for example, how to work, how much to work, what the right quality is, etc. These norms are socially sanctioned within because the shirking agent is not just hurting himself; rather he is hurting the entire group with his behavior. Hence, the employee who obeys the social norms will not only gain recognition from her fellow group members she will also be rewarded through the organization’s incentive system. In contrast with the results produced by the incentive system discussed in the Hawthorne studies, Coleman (1993b) seems here to arguing that the constitution matters, if such an incentive system is to work, because if the organization is designed in a manner which allows

norms to arise in accordance with the organization's goals, then these will increase the frequency of the desired behavior. In other words, the organization needs to be designed in a manner which allows social extrinsic incentives to be aligned with economic incentives. In such an organization no agent/principal problem will emerge because the agent's desires, preferences, and beliefs will be aligned with the social norms which are aligned with the organization's goals. The functional dimension of Coleman's (1990; 1993b) social norm theory is here clearly illustrated in the sense that these social norms serves a particular purpose. Moreover, by aligning the economic and social incentives within the organization the distinction between the formal and informal social system also disappears at least in the conventional antagonistic sense which have been inherent to the discipline, at least, since Barnard's (1938/1968: 116) discussion on the matter. In some way it can be argued that the motivational gap addressed by Coleman (1993b) is the same as Barnard (1938/1968: 89) addressed, the motivational gap between the organization's goal and the employee's motives – a focus which makes both authors less observant of the extrinsic/intrinsic motivation permeating the current discussions (see the next chapter). Furthermore, like Barnard (1938/1968: 42) he seems to be arguing that the foundation of an efficient organization is; *firstly* to get individuals to cooperate; and *secondly* to get the group to cooperate in a manner which is favorable to the organization. Unlike Barnard (1938/1968: 57), however, he does not seem to argue that the social agents are motivated by vertical appraisals from their managers; rather they are motivated by social approval from their peers and, of course, economic incentives.

Hobbes' (1651/1985) and Coleman's (1990) discussion of the exchange of rights draws the analytical focus towards the structuring of social interaction, and in particular the relational nature of interaction and the values on which these are centered. According to Hume (1739/1985: 535) the concept of justice is not derived from nature, but is artificially constructed, and cultivated through education and socialization. In contrast to Hobbes, Hume distinguishes between desire and gain, whilst social agents are motivated by gains, to restrict these by means of a stable distribution through social norms is not opposite desire, rather it is a matter of cultivating the social element in man (Hume, 1739/1985:541). Hume writes further on this matter:

“By this means, every one knows what he may safely possess; and the passions are restrain'd in their partial and contradictory motions. Nor is such a

restraint contrary to these passions; for if so, it cou'd never be enter'd into, nor maintain'd; but it is only contrary to their heedless and impetuous movement. Instead of departing from our interest, or from that our nearest friends, by abstaining from the possessions of others, we cannot better consult both these interests, than by such convention; because it is to their well-being and subsistence, as well as to our own." (Hume, 1739/1985:541)

What Hume is arguing here is that insofar as the rational agent is driven by desire (from an internalist perspective) it would make no sense if she was to abstain from this, because what would the impetus then be to enter into such a social state. Such a convention which the social and rational agent adheres to is however not a promise or an agreement, rather a product of strategic reasoning, that is, it is centered upon a common interest and based upon the actions of the other, which implies that as long as the other social agent upholds his end, the agent will also uphold hers (Hume, 1739/1985: 541-542). Hume argues:

'Tis certain, that no affection of the human mind has both a sufficient force, and a proper direction to counter-balance the love of gain, and render men fit members of society, by making them abstain from the possessions of others. Benevolence to strangers is too weak for this purpose; and as to the other passions, they rather inflame this avidity, when we observe, that the larger our possessions are, the more ability we have of gratifying all our appetites. There is no passion, therefore, capable of controlling the interested affection, but the very affection itself, by an alteration of its direction. Now this alteration must necessarily take place upon the least reflection; since 'tis evident, that the passion is much better satisfy'd by its restraint, than by its liberty, and that in preserving society, we make much greater advances in the acquiring possessions, than in the solitary and forlorn condition, which must follow upon violence and an universal licence. (Hume, 1739/1985: 543-544)

In that sense, the social convention is based upon interest, and both parties will solely adhere to this as long as it is within their private interest to do so (Hume, 1739/1985:550). Although the initial motive behind justice is self-interest, Hume (1739/1985: 551) adds to this, what motivates socially rational agents to adhere to these is sympathy – the emotional foundation of morality – for one's peers. In other words, violating these social conventions not only causes anticipated disutility to the norm violator stemming from the future state which it might lead to, but also, emotionally, from the violation of the human bound based on sympathy. To some extent this resembles Smith's (1759/2009) idea on social rationality, albeit Smith did not take the utility dimension into account, at least not the utility of the violator; rather the violator's disutili-

ty derived from her taking the perspective of the other and reasoning about how her actions might harm the other (see the section on Adam Smith below). The logic of this social emotion of sympathy (or empathy it might be called today) seems to be founded on an idea of human respect in the case of Hume and respect combined with understanding in the case of the perspective derived from Smith.

Besides Hume and Smith, the French philosopher Jean-Jacques Rousseau formulated a similar idea of human bounds as being based on the sentimental foundation of sympathy, although in his understanding it was more to resemble with pity. Rousseau wrote “A Discourse on Inequality” (1755) a work which was a critique of Hobbes’ natural philosophy. Like Smith, Rousseau argued that social interaction was not entirely based on rationality, albeit he distinguished between the *natural* social sentiment (sociality) pity or compassion, and those social sentiments stemming from being socialized into a distinct set of moral rules (sociability) (Rousseau, 1755/1984: 70). While Hobbes’ may be viewed as the first to identify the social level, Rousseau may be viewed as the first to identify, how social behavior constituted social institutions, and how these social constructions came to govern social life (Rousseau, 1755/1984: 70, 13, 114).

Rousseau (1755/1984: 70, 86-87, 89, 90, 98) describes the human (physical man) in the initial position as a simple, solitary human, motivated by self-preservation and, although, he might had had no sense of good or evil because he was not yet bound by any moral code, he felt pity (moral metaphysics) toward those he encountered, he had no desires beyond his basic physical needs, and as such he was not prone to long for future states, or compare his present state to that of others. Rousseau is here describing a kind of basic and non-reflexive social sentiment not yet deprived of its purity by moral codes enforced and governed by those powerful enough to do so (Rousseau, 1755/1984: 99, 101). Rousseau defines the social role of this sentiment as:

“It is therefore very certain that pity is a natural sentiment which, by moderating in each individual the activity of self-love, contributes to the mutual preservation of the whole species. It is pity which carries us without reflection to the aid of those we see suffering; it is pity which in the state of nature takes the place of laws, morals and virtues, with the added advantage that no one there is tempted to disobey its gentle voice; it will always dissuade a robust savage from robbing a weak child or a sick old man of his hard-won sustenance if he has hope of finding his own elsewhere.” (Rousseau, 1755/1984:101)

Rousseau argues further:

“Do good to yourself with as little possible harm to others. In a word, it is to this natural feeling, rather than to subtle arguments, that we must look for the origin of that repugnance which every man would feel against doing evil, even independently of the maxims of education.” (Rousseau, 1755/1984:101)

As such the human in the initial position was no brute who needed a social contract to bind him nor did he need a moral code transcending the natural social sentiment. Contrary to Hobbes’ perspective, social order needs not necessarily to be founded on a common law or abstract reason and calculation besides the one prescribed by nature (or instinct). In that sense, what Rousseau is skeptical towards is the assumption on rationality described by Hobbes – that is, the assumption of egoistic rationality, and the assumption of social rationality as being governed by the social contract. Furthermore, sociability and the need for moral codes arose simultaneously with vanity, shame, and envy (Rousseau, 1755/1984: 114). According to Rousseau (1755/1984: 114, 115, 116, 117, 119), this was due to the institutionalization of property and the differentiation and competition which followed from this, combined with the emergence of new social arrangements, the creation of communities, which enabled comparison of abstract qualitative dimensions such as wealth, beauty, and admiration. That is, *social preferences* originate from the social creation of inequality as Rousseau seems to argue in the following quotation:

“From those first preferences there arose, on the one side, vanity and scorn, on the other, shame and envy, and the fermentation produced by these new leavens finally produced compounds fatal to happiness and innocence.” (Rousseau, 1755/1984: 114)

In that sense, the modern individual, as described by Rousseau (1755/1984: 119), cares about her social standing because she is constantly confronted with differences, just as she is socialized into a social symbolic world in which different external items, for example symbols on wealth or human actions, has taken on social symbolic value – a value which differs from the functional value. Here it seems that Rousseau is in line with the perspectives of both Hobbes and Hume.

Rousseau, then, argues against reason and rationality as the foundation of morality and prefers an emotional foundation. Rousseau is, compared to Smith, more radical in

his argument which, to some extent at least, is because he romanticized the natural human. Smith, on the other hand, did not share Rousseau's critical notion of modern society but viewed socialization as a way of cultivating the basic social sentiment, hence his understanding of social rationality as sympathy and followed by virtues. Of course, this is also a meta-ethical debate between, on the one hand, rationality (cognitive) and emotivism (non-cognitive), and, on the other hand, between nature (instinct) and nurture (socialization).

Hume and Rousseau, then, it seems, through their critique of Hobbes, adds a further dimension to the nature of aligning the social norms with economic norms insofar as they bring into the relation, social emotions. Instead of reducing acceptance to a matter of rational self-interest, these authors argue that social emotions matter and in particular sympathy which I will further discuss in the next section. While Rousseau's perspective romanticizes the initial position and as such may only be viewed as a periphery critique of Hobbes' conception, it does bring into the discussion an important perspective – namely that social agents hold social preferences (see also chapter V), and that the violation of this harmonious state of non-social life characterizing the initial position necessitates the installment of social institutions governing social life. That is, norms on resource distributions may emerge to support social interaction between differentiated social agents. Coleman's discussion of social norms and his discussion of these as support of formal structures seem, at least to some extent, to support Lawler's conception. However, might this also be the case if the motive which enables this is not rational egoism, but sympathy? That is, if sympathy is the primary guard against egoism and injustice, and as such the foundation of fairness, a system which substitutes sympathy with technical matters might undermine the perceived fairness of the system insofar as the disutility may exceed the utility in the sense that the social agent who violates some social norm governing horizontal relations might come to view herself as being disloyal to her colleagues. While both Coleman and Hume emphasize the importance of solidarity as the foundation of group relations, they also reduce it to a matter of disutility either as costs or as emotional displeasure. Smith as I will discuss in the following section puts forward a different conception.

### 3.3.3 Adam Smith

Plato's account of rationality was centered upon the governing of the self and upon obtaining true desires and as such no *true* desire was against reason – the role of rationality was to secure a balanced living. In accordance with Plato's view, self-interested actions meant nothing but the fact that individuals act in their own interest. The modern usage of self-interest as egoism is often associated with the perspective developed by Hobbes, even though he, as argued above, identified the individual and social limits of egoism. Nevertheless, the rational individual, identified in Hobbes' philosophy, is a *limited* egoist, always acting in his own interest and with his own preservation as the primary rule of reason. The constitution of social order was then based on this particular kind of semi-strong rationality, limited egoism, and a social system based on rational laws. A similar logic of constitution is found in Adam Smith's work "The Wealth of Nations" (1776). Smith addresses in his work the constitution of the economic system and not, as such, the political constitution. Nevertheless, his philosophical work follows the lines created by Hobbes before him.

The theory on rationality that Smith proposes in his two works "The Wealth of Nations" (1776) and "The Theory of Moral Sentiments" (1756) can be viewed as an early statement of the utilitarian perspective, although the first complete doctrine was made by Jeremy Bentham (1780). Smith distinguishes in his philosophy between economic and moral philosophy, thus he develops on Hobbes' argument on the limits of egoism, and distinguishes between economic rationality and social (moral) rationality.

Smith's theory on economic rationality is presented in his work "The Wealth of Nations". Contrary to Hobbes, Smith argued that cooperation between individuals was not founded on any social contract, but rather "*...the accidental concurrence of their passions in the same object at that particular time.*" (Smith, 1776/2003: 23). What Smith is arguing here is that cooperation between individuals emerges from their self-interested pursuit of their own wellbeing – what Oliver Williamson (1996:30) calls the spontaneous government of the market. According to Smith (1776/2003: 10), the economic system is built upon the division of labor, thus the relations between individuals is centered upon exchange of goods and services. The exchange relation between economic agents differs from exchange relations between social agents because, while the

former is based on self-interest, the latter is based on benevolence (Smith, 1776/2003: 23). Smith argues further on the exchange relations between economic agents:

“Give me that which I want, and you shall have this which you want, is the meaning of every such offer; and it is in this manner that we obtain from one another the far greater part of those good offices which we stand in need of. It is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own interest. We address ourselves, not to their humanity but to their own self-love, and never talk to of own necessities but of their advantages.” (Smith, 1776/2003: 23-24)

The economic rational agent in economic exchange relations is, then, solely motivated by her own self-interest and as such has no obligation towards the other agent which extent the economic exchange itself. In other words, the economic rational agent always asks “what is in it for me”, thus she only takes part in an exchange if it is advantageous to her, and as such she has no principles governing her actions, except from those which govern the market. The non-rational economic agent is, then, an agent who takes part in an exchange even if it is disadvantageous to her, she is motivated by altruism, and she believes other agents also to be so. While it seems that the economic rational agent is build incomplete on pure self-love, it seems that the economic non-rational agent is an irrational fool most likely to stand prey at the mercy of other agents. However, it is important to note that Smith solely associates this particular behavior with economic relations, and by dividing his philosophical system into an economic and moral philosophy he explicates not only the two different kinds of rationality associated with social and economic relations, but also the fact that humans are motivated by social-interested as well as self-interested actions.<sup>20</sup> As such he moves beyond the moral philosophy of Hobbes’ by arguing that the foundation of morality is not self-preservation but social interest.

---

<sup>20</sup> This distinction between social and economic rationality which Smith created and which was by J.S. Mills (1874/2000: 89; 93) turned into a distinction between political economy and ethics (notable Kantian ethics) – that is, a functional division of scientific subjects, entailed a division of moral and economic philosophy.

### **3.3.3.1 Sociality and Rationality**

Smith argues in his moral philosophy that morality is centered upon and, as such, can be reduced to, the social emotion of sympathy:

“How selfish soever man may be supposed, there are evidently some principle in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it except the pleasure of seeing it.” (Smith, 1759/2009:4)

Contrary to the economic rational agent, the social agent cares about the wellbeing of his peers because it affects his own wellbeing by a grip on his emotions. In that sense, the utility of some object is only affected by the utility it brings the agent, in hedonistic terms – pleasure, but also how this affects his fellow agents. This fellow-feeling, as Smith (1759/2009:5) calls it, is based on introspection and imagination, that is, the acting agent must consider how his action would affect the other agent by trying to imagine how he would feel if this particular action was taken against him (Smith, 1759/2009: 4). Smith is, of course, aware of the limit of this method, because one agent can never, as such, put himself in the position of the other – that is, he can never feel as the other would feel. Emotions are here used as the foundation of utility because they take on hedonic qualities – some external objects invoke pleasure while others pain. Additionally, emotions also take on a normative element prescribing not only how one should act but also how one should feel and the intensity of this feeling, that is, it relates to situations of emotional dissonance (Smith, 1759/2009: 11, 12, 22). Emotional dissonance occurs in situations when the emotional display of the agent is contrary to the emotional display of the other social agents, or perhaps when the agent believes that he should feel *A* but actually feels *B*. The social rational agent, then, holds beliefs on which emotions are appropriate in which situations and the intensity of these emotions. These beliefs might either be created through introspection, as argued above, or by the existence of social institutions (Smith, 1759/2009: 28).

Smith (1759/2009: 14), then, uses sympathy in a double sense; on the one hand, the social agent feels sympathy towards another agent; and on the other hand, he is sympathetic towards the judgments of this other agent if he feels an emotional correspondence. In juxtaposition to Plato’s and Hobbes’ perspectives, it seems that the principle of reason is substituted by a principle of correspondence, that is, an action is

judged as irrational, in the eyes of the other, if it does not correspond to how the other agent, in a similar position, would act (Smith, 1759/2009: 16). On the matter of this assessment, Smith argues that the observer assesses both the cause and the consequences of these actions to the acting individual, that is: firstly, does the response correspond to the object, or perhaps correspond to how the observer himself would respond – is the response appropriate; and secondly, is the response advantageous to the acting agent (Smith, 1759/2009: 13). Propriety, he argues, as the principle of utility, affects the process of choice by its grip on the emotions and subsequently its effect on the social utility of some end (Smith, 1759/2009: 169).

“In the steadiness of his industry and frugality, in his steadily sacrificing the ease and enjoyment of the present moment for the probable expectation of the still greater ease and enjoyment of a more distant but more lasting period of time, the prudent man is always both supported and rewarded by the entire approbation of the impartial spectator, and of the representative of the impartial spectator, the man within the breast.” (Smith, 1759/2009: 194)

This approach to social rationality is, then, not, entirely that is, based on selfish self-government, but also on the emotional utility of sympathy. The socially rational agent’s utility is positively connected to the utility of the other social agents, and he is motivated by social-interested actions because he feels sympathy towards them and is admired for this social virtue. On the other hand, the social individual who constrains himself is motivated by the selfish desire for dignity and honor (Smith, 1759/2009: 18, 19). Smith argues on the result of this social motivation:

“And hence it is, that to feel much for others and little for ourselves, that to restrain our selfish, and to indulge our benevolent affections, constitutes the perfection of human nature; and can alone produce among mankind that harmony of sentiments and passions in which consist their whole grace and propriety. As to love our neighbor as we love ourselves is the great love of Christianity, so it is the great precept of nature to love ourselves only as we love our neighbor, or what comes to the same thing, as our neighbor is capable of loving us.” (Smith, 1759/2009: 19)

The individual governed by his own principles and motivated by his own honor is as such motivated exclusively by a selfish desire not to lose face – he wants to avoid shame. The socially motivated individual is, as such, also governed by his sense of “the eyes of the others” however, not because he wants to avoid shame, but because he feels sympathy. In that sense, the social motivated individual does not refrain from pursuing an end

because it is disadvantageous to him, but because it might hurt some other agent, and because of this social connection built upon sympathy it also hurts the acting agent. This perspective, then, goes beyond a cognitive belief on consequences combined with self-preservation, towards a non-cognitive understanding based on introspection, emotions, and social sentiments - perhaps better understood as the social norm which governs a social relation. Smith's position is, then, centered between consequentialism and deontology. For example, if Alter is a consequentialist he would refrain from lying because he fears getting caught; while if he was a deontologist he would refrain from lying because he holds cognitive principles based on the categorical imperative, that is, he does not lie because lying, if elevated to a social value, would be socially undesirable; the social rational individual who believes Smith was right, refrains from lying because it would hurt the target of this deceiving conduct, and as such it produces within the individual the equivalent feeling of how he would feel if someone was to lie to him. In other words, the beliefs on what is right or wrong is based on the agent being socialized into the norms and customs of society, and they present themselves not as cognitive recognized reasons, but as non-cognitive anticipation of emotions it will produce – in short, the social rational agent does not lie because it makes her feel bad.<sup>21</sup>

Smith, however, did not take the position that the social agent was either motivated by self-interest or social-interest. Rather, as already argued, he claimed that social agents are equipped with both motives. Prudence, Smith argues, is the foundation of rationality. Prudence is the capability of the agent to produce actions which are, not only, causally related to a particular end which the agent wants to bring around, but also the ability to produce agreeable actions, that is, by the introspective process of taking the perspective of the other (Smith, 1759/2009: 192). As such, this capability, learned through experience, is both procedural and normative in nature. The rational agent cares not solely about bringing about some external material ends, for example wealth, but also about her social stand, the respect she is ascribed by her peers (Smith, 1759/2009: 192). Smith seems here to be inspired by Hobbes' notion of the human nature, and in

---

<sup>21</sup> What Smith seems also to capture is the complex relation between self-interest and social interest, and the fact that self-interested and social-interested actions sometimes produces similar results. Hobbes' escaped this problem by considering all ends and actions as self-interested.

particular two social values – competition and glory. The rational social agent, Smith argues contrary to Hobbes, is not only trying to preserve her own life and pursuing her selfish material ends in accordance with this principle; the rational agent also cares about her relative stand in society. That is, her utility is, as such, not completely that is, depending upon the end wants to bring about, but also wealth and fortune of those around her. Above the principle of social comparison defined and related to the evaluative dimension of rationality, but also as related to the assessment of one's results compared to those with who the individual compares herself.

Moreover, Smith also argues that this status motive generates an emotional bias, because “[w]e suffer more (...) when we fall from a better to a worse situation, than we ever enjoy when we rise from a worse to a better.” (Smith, 1759/2009: 192). That is, the individual who fall from a better to a worse situation often associates this with a loss of status, and this can either occur because the individual herself fall or because those around her rise. The rational individual, then, want to maintain status quo, that is, avoid the risk status loss, just as she wants to avoid the risks of not being able to secure a continuous fulfillment of her desires (Smith, 1759/2009: 192). In other words, the rational agent is risk averse in regard to situations which may affect either her ability to future procurement or her social stand. Smith argues:

“He has no anxiety to change so comfortable a situation, and does not go in quest of new enterprises and adventures, which might endanger, but could not well increase, the secure tranquility which he actually enjoys. If he enters in to any new projects or enterprises they are likely to be well concerted and well prepared. He can never be hurried or drove into them by any necessity, but has always time and leisure to deliberate soberly and coolly concerning what are likely to be their consequences.” (Smith, 1759/2009: 194)

The rational agent prefers security to risk, and she possesses an ability to govern her desires which enable her not to give into immediate pleasures, but aim for those long term pleasures, thus she has an ability to treat those long term pleasures as they were present (Smith, 1759/2009: 170). The rational agent bases her decisions on “...*real knowledge and skill in [her] trade or profession...*” (Smith, 1759/2009: 192), just as she engages in careful deliberation before she acts. As such, it may be argued, that Smith does not push his understanding of rationality by that of Hobbes and Hume, his understanding is still build on rational actions as motivated by self-interest, albeit his princi-

ples of social utility as a combination of utility and propriety adds an important social dimension to the matter. However, Smith does in fact try to push his understanding further than those before him when he argues that even though self-interested rationality, prudence, is admirable in humans it is a cold admiration not to be confused with the warm glow of social virtues (Smith, 1759/2009: 195). Smith argues further:

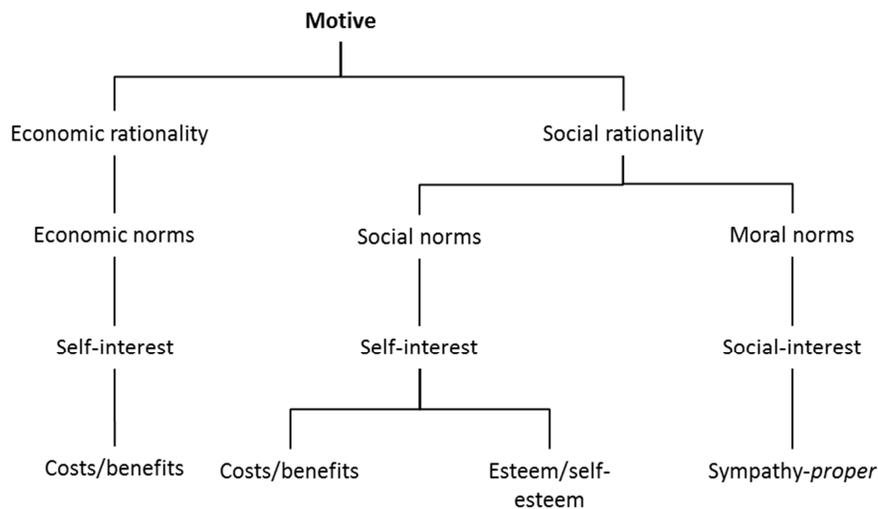
“Wise and judicious conduct, when directed to greater and nobler purposes than the care of the health, the fortune, the rank and reputation of the individual, is frequently and very properly called prudence. We talk of the prudence of the great general, of the great statesman, of the great legislator. Prudence is, in all these cases, combined with many greater and more splendid virtues, with valor, with extensive and strong benevolence, with a sacred regard of the rules of justice, and all these supported by a proper degree of self-command. This superior prudence, when carried to the highest degree of perfection, necessarily supposes the art, the talent, and the habit or disposition of acting with the most perfect propriety in every possible circumstance and situation. It necessarily supposes the utmost perfection of all the intellectual and of all the moral virtues. It is the best head joined to the best heart.” (Smith, 1759/2009: 195)

The distinction between instrumental rationality and some stronger kind of rationality becomes here centered upon a distinction between self-interested rationality and social-interested rationality which also takes into account moral principles, that is, rationality in this form demands a knowledge of right and wrong actions, just as it demands a knowledge of morally good and morally bad actions. This particular form of rationality demands that the agent is socialized into a particular society, and as such, in matters of judgment she takes into consideration the anticipated judgments of her peers, and applies these as means to self-government. The point of departure here is not solely the abstract or practical reason, but also her cognitive schemes generated partly by socialization and partly by experience.

What seems interesting here, is that Smith recognizes the impact of social institutions on individual rationality, just as he recognized their impact on the economic agent and the rationalization of exchanges as organized by the existence of a market, trust, price mechanisms, and money. Contrary to the self-interested rational agent, the rational social agent is bound to other agents through the social feeling of mutual empathy – that is, he feels sympathy towards his peers, which entails a particular kind of social-interest, just as he understands his own actions, and those of his peers through this sympathetic

social connection as being in accordance with or contrary to the governing social institutions. Plato applied the distinction right and wrong actions, while Hobbes based his distinction self-preservation at the individual and social level, and generated a distinction between self-preserving egoism and madness (at the level of the individual) or war (at the social level). Smith (1759/2009: 195) argues that self-interested rationality is centered upon a distinction between advantageous and disadvantageous actions, while social rationality is built upon a distinction between being in accordance with the governing social institutions or not; and self-interest combined with other egoistic vices, for example envy or glory, as opposite to social interest based on sympathy combined with other social virtues, for example justice or indignation. This might be illustrated as follows:

Figure 7: Hierarchical taxonomy of motives according to Smith.



According to the perspective developed by Smith, then, one can distinguish between economic rationality inherent to economic exchanges and based solely on instrumental rationality; and social rationality which might either be centered upon self-interested motives or social motives. Self-interested social rationality can, furthermore, be divided into an instrumentalist approach associated with Hobbes (and Coleman), and a non-rational esteem/self-esteem approach associated with Hume. Social rationality motivated by social-interest can be perceived as sympathy-proper in the sense that it is centered upon a disinterested other-regarding motive. This latter category can also be associated with the Kantian approach (see the section Immanuel Kant). Moreover, Smith's distinc-

tions also enable a distinction between different sets of normative expectations which might govern social interaction. This, however, should not be understood as mutually exclusive sets of normative expectations insofar as these may be mixed at some occasions – that is, economic transactions are also governed by social and moral normative expectations, just as moral action might also be influenced by economic concerns or concerns about esteem and self-esteem – see also chapter IV of this thesis for a further discussion. However, this spill-over from one domain to another may cause irrationality to emerge, because, as Smith (1776/2003) seems to argue, the economic agent who appeals to the moral values of another economic agent may be perceived as economic irrational. However, viewing irrationality as deviations from normative standards may not be perceived as irrationality-proper, as I will argue in the section below on Jon Elster and Donald Davidson; rather they can be perceived as non-rational actions – that is, actions based on motives emerging from a different set of normative expectations. I will in chapter VIII return to the discussion between social and moral norms. Furthermore, the conceptual program developed by Smith also captures the central distinction on which the modern conception of fairness is based – discussed above – insofar as it might be argued that the instrumental perspective is based on a social rationality based on social norms which focus on the costs incurred by the agent in her pursuit of gaining control, whereas the relational perspective is closely related with the Humean perspective developed above; and Folger’s perspective can be categorized as belonging to a tradition which views fairness as based on moral norms – as such, the social rationality of agent is governed by morality and not by social circumstances related with, for example, sanctions or esteem/self-esteem.

Although Lawler does not neglect that social agents may be motivated by social needs, he does argue, along with Coleman, that this can be reduced to a matter of cost/benefits. That is, rational agents will pursue whatever might be within their interest as a matter of a cost/benefits analysis. Moreover, as argued in the previous chapter esteem/self-esteem concerns might also be aligned with economic interest to generate a social symbolic foundation of economic incentives. However, these esteem/self-esteem concerns also have a horizontal dimension related to the employee-colleague relation and the esteem gained vertically through a positive performance appraisal may be diminished if cannot be exchanged into horizontally recognized symbols, as also seems to

be Lysgaard's (1961/2001) basic argument. The motives behind fairness concerns may, according to this perspective then, be divided into three basic kinds, economically, socially, and morally. That is, the question whether something is fair or not has, at least, three dimensions. The technical assessment can, primarily, be associated with the economic dimension, while the social dimension may be viewed as purely horizontal, and the moral dimension may be viewed as a mixture. In other words, rational acceptance is more than economic rational acceptance, it is also social rational acceptance, and as such it is related with a social and moral dimension not contained in the purely technical approach.

### **3.3.3.2 *The Impartial Spectator and Public Sensitivity***

Raymond Boudon (2001) argues in opposition to Coleman's (1990) perspective that matters regarding solidarity cannot be reduced to a cost/benefits analysis. His theory is based on Smith's (1759/2009) idea about the impartial spectator:

“[W]e are pleased to think that we have rendered ourselves the natural objects of approbation, . . . and we are mortified to reflect that we have just merited the blame of those we live with.” (Smith, 1759/2009:138)

He goes on later in the text:

“We endeavor to examine our own conduct as we imagine any other fair and impartial spectator would examine it. If, upon placing ourselves into the situation, we thoroughly enter into all the passions and motives which influenced it, we approve of it, by sympathy with the approbation of this supposed equitable judge. If otherwise, we enter his disapprobation, and condemn it.” (1759/2009: 133)

What Smith is here arguing, it seems, is that social opinion works through the social agents' need for esteem and self-esteem. By taking the role of the *impartial spectator* the agent views herself from the perspective of the other – that is, in a sense she is objectifying herself through the perspective of social norms, customs, etc. Boudon (2001) defines the effect of the impartial spectator as follows:

“...it designates a complex process by which individual opinions and judgments, biased by the interests and passions of those involved, *may*, under some conditions, in the aggregate, produce an opinion or a judgment in keeping with the common interest.” (Boudon, 2001: 182, italics in original)

In Boudon's (2001) perspective, then, this impartial spectator represents the *otherness* which inhabits social agents' life insofar as it represents interests and passions which are not inherent to the agent self. Coleman's (1990) perspective may also be understood along this line of reasoning because he argues that the transfer of the right to influence occurs across temporal and spatial frames – that is, these relational structures which emerges from the transfer of rights cannot be reduced to the present moment or the immediate present social agents, inasmuch as one can transfer these rights to agents encountered in the past, agents living in other countries, strangers, or non-human agents, for example, social institutions, religion, etc. Moreover, as Boudon (2001:185, 187, 211) argues, it is this *immanent otherness* which enables social agents to pass non-selfish judgments in the sense that it is the ability to address general values and set aside one's own desires which prevails and dominates in the short run, in what he calls the partial spectator or the engaged agent living in the present. In organizations, for example, this might occur when the employee takes decisions on some matters which reflect not only her own self-interest, but also the non-selfish interest accompanying her *role* in the sense that, for example, an employee working in a public organization is able pass judgments which not only reflects her private interests, but also what she believes to be best for the client – I return to this matter below in my discussion of Public Service Motivation (see chapter IV). Boudon (2001) argues on this matter:

“...some interests may include both a particular dimension and a general dimension, or to put it otherwise, an egoistical and an altruistic dimension, in which altruism is not simply a more intelligent form of egoism.” (Boudon, 2001: 187)

To some extent it might be argued, as Coleman (1990) did, that social agents adapt their opinions to the public opinion due to different bias, for example, because they perceive it to be within their interest, due to role concerns, or due to emotional and affiliative concerns (Boudon, 2001: 189, 191, 193, 201). However, this social agent should not be understood as one who *just* accepts public opinion and adopt it as her own; rather the social agent has reasons to believe that it is right – that is, these beliefs also have a cognitive aspect in the sense that some belief is held to be true because it is associated with a theory held by the agent (Boudon, 2001: 195, 210; see also the section on Raymond Boudon below). The effect of the impartial spectator, then, may explain why certain opinions are held in the sense that it illustrates the linkage between private desires and

collective values, and as such it illustrates how the partial spectator anticipates the influence of the impartial spectator when she passes a judgment on a particular subject matter by invoking some collectively held value – in a way this impartial spectator illustrates the agent’s ability to perform judgments of a transcendental nature.

From what has been argued here, it seems to follow that social life cannot be reduced to purely instrumental or strategic reasoning about how one best brings about some change which is in accordance with one’s ends, insofar as agents are also emotionally connected with one another. The emotional foundation of social rationality, identified above by Smith is different from the ones associated with the writings of Hume and Rousseau, in the sense that individuals do not just feel pity towards one another nor do they solely care about one another due to selfish concerns about esteem/self-esteem. The latter notion also emphasizes the connection between: “how I relate to myself” and “how others relate to me”, just as it seems to introduce a mediating form: “how I relate to others”. Social influence one another and they care about other social agents’ opinions – that is to say, they do not live in a social vacuum. Through this sense of public opinion, it becomes possible for the social agents to reflect upon how other social agents might react upon their actions – how they might perceive them. In a sense then, it is through this that one is capable of taking the perspective of the other – to transcend one’s immediate self. The incentive being partly selfish and self-directed; and partly other-directed, in the sense that one shares the *pain* one inflicts upon other individuals. Social rationality in this sense, then, requires more than sensitivity to governing social norms - it presupposes sensitivity towards other social agents – insofar as navigation with a social context also implies navigation among other social agents.

### **3.3.4 David Hume**

In this section I will discuss Hume’s approach to rationality. The aim of this section is primarily to introduce Hume’s conception as a contrast to the Platonic perspective. This contrast becomes important as a link to the next section on Kant insofar as his perspective can be viewed as being in opposition to Hume’s perspective.

From what has been discussed so far in this text, it seems that Plato was concerned with the motivational conflicts which emerges from the three parts of the soul,

and how this conflict could be resolved by the imposition of a second-order reason, and a third order desire. The basic motivation in Plato's philosophy is the pursuit of the good, in a hedonic sense that is. The rational agent using the power of her reason can separate what is truly good, from what only appears so. Plato did not, probably due to his principle of non-contradiction (the fact that opposite desires cannot exist simultaneously), consider ambivalence as a matter of motivational conflicts, but would probably argue that the sides would need to be united under a common constitution, or perhaps that the lesser desire would be suppressed (depending on which psychological principle applied in the dialogue). That is, there is always some end and some choice which presents itself as a better one, whether this selection is based on reason, spirit or desire is another matter. Additionally, the doubt felt by the agent could be removed if she used her power of reason, because this would provide her with an accurate assessment of how best to act, and which ends to follow. Or perhaps, if his idea of unity is taken into consideration, the agent should listen to herself. Hume, however, considered the Platonic distinction between reason and passion to be false, because, even though, the mechanical procedures of mathematics are useful, they cannot make the agent act, insofar as they only affect the practical beliefs concerning the object, and not the evaluative dimension. Hume argues:

“Abstract or demonstrative reasoning, therefore, never influences any of our actions, but only as it directs our judgment concerning causes and effects; which leads us to the second operation of the understanding.”(Hume, 1739/1985: 461)

Hume, then, identifies the role of reason as instrumental and based on beliefs about causality. Hume argues on this matter:

“’Tis obvious, that when we have the prospect of pain or pleasure from an object, we feel a consequent emotion of aversion or propensity, and are carry'd to avoid or embrace what will give us this uneasiness or satisfaction. 'Tis also obvious, that this emotion rests not here, but making us cast our view on every side, comprehends whatever objects are connected with its original one by the relation of cause and effect. Here then reasoning takes place to discover this relation; and according as our reasoning varies, our actions receive a subsequent variation. **But 'tis evident in this case that the impulse arises not from reason, but is only directed by it.**” (Hume, 1739/1985: 461; my emphasis)

That is, some objects bring pleasure while other pain, as such the agents are drawn toward those which bring them pleasure, and repulsed by those which bring pain (Hume, 1739/1985: 167). The object of these causal beliefs is to establish a cognitive link between the hedonic experience and the object based on reason and experience. In that sense, beliefs should be understood as having a mediating role and acting on the will through its *preferences*, as the presence of *past* impressions (temporally bound to the present and particular in nature) – what Hume called ideas (temporally bound to the past and universal in nature), although one should recognize that beliefs, in a sense, have a stronger impact on the mind than a mere idea seems to have, an impact which resembles that of a present impression (Hume, 1739/1985: 51-52; 168). In that sense, the beliefs work on the passions, thus transform these into preferences – or more perhaps more accurate, belief-induced desires, see also figure 4 above (Baier, 1994: 166). Moreover, this particular conception of beliefs also allows the agent to be motivated by the presence of a *past* impression, and as such it allows the agent to anticipate the future consequences of her actions, thus making it possible for her to avoid a future pain, just as it makes pursue certain ends because they are associated with pleasure. In other words, agents hold beliefs on causal relations and from these causal relations floats the desire to a particular end, and subsequently a motive to take on the necessary actions needed to bring this end around (Korsgaard, 2009: 60). The agent then, who holds a particular belief regarding a particular means-end relation will, as such, also be motivated to take these means solely because they are causally necessary in regard to this particular means-end relation. Consider, now, Hume’s famous restatement of the Hobbesian argument:

“Thus it appears, that the principle, which opposes our passion, cannot be the same with reason, and is only called so in an improper sense. We speak not strictly and philosophically when we talk of the combat of passion and of reason. Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them.”(Hume, 1739/1985: 462)

Hume is here arguing against the combat model of the soul, however, not as Plato did because it required agents without essence, but because it is built on a false dichotomy. One might argue that he escapes the problem himself by removing the distinction. Nonetheless, reason alone cannot be the ruler. The domain of reason, true or false, is too

weak and it cannot produce any actions, only indirectly by guiding these towards the end determined by passions, and as mediating elements working on the will (Hume, 1739/1985: 510). To some extent at least, this resembles Plato's argument on the limited motivational effect of reason alone without being aided by thymos. In that sense, as I argued above, there is a practical distinction between knowing something and wanting something.

In the above quotation on the function of belief, Hume also seems to argue that when (or if) reason opposes the motivation stemming from the passions it is because it is misguided by its causal beliefs. That is, the agent who is judged as irrational is not judged as such, because she follows some *irrational* end, but solely because she is guided by false beliefs (Baier, 1994: 164). This is still today, more than two hundred years later, a controversial idea, and it becomes even more controversial if one takes the following quotation into consideration:

“Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger. 'Tis not contrary to reason for me to chuse my total ruin, to prevent the least uneasiness of an *Indian* or person wholly unknown to me.” (Hume, 1739/1985: 463)

As Hume seems to argue in this quotation, it is not opposite to reason to be a pure altruist, nor is it contrary to reason to prefer the destruction of the world to a minor bodily discomfort. Insofar, as reason does not generate any desires at all, it cannot, as such, be opposite to a desire. According to this understanding, it would not be opposite to reason, should the agent who is trying to stop smoking, smoke a cigarette, because reason solely informs her that she is sacrificing her future benefits by following this act through (Baier, 1994: 165). This becomes even clearer if the continuing lines of the quotation are taken into consideration:

“'Tis as little contrary to reason to prefer even my own acknowledge'd lesser good to my greater, and have a more ardent affection for the former than the latter.” (Hume, 1739/1985: 463)

In that sense, this agent may be judged as weak-willed but not as being irrational. This particular conception also makes Hume (1739/1985: 509) reject the idea that the foundation of morality could be reason, because morality is about exiting or preventing a

particular action, hence its foundation must necessarily be the passions, insofar as these are the only entities possessing such capacities.

One may, of course, question this logic, because if the only way an agent can be instrumentally irrational is if she fails to take on the necessary means to bring her desired end around, then, the distinction between rationality and irrationality seems to be in danger of becoming tautological. That is, if an agent did not succeed in bringing a particular end around the agent could always claim that she was not trying to do that – she did not consider it as her end – and as such she would not be irrational (Korsgaard, 2008: 39). Consider Hume’s own statement on this matter:

“The moment we perceive the falsehood of any supposition, or the insufficiency of any means our passions yield to our reason without any opposition.” (Hume, 1739/1985: 464)

Another example might be, if an agent is trying to put a screw into a wall by means of a hammer – he should not be considered as irrational if he holds a belief stating this as the proper way. In that sense, if nothing can be considered as irrational, then, by mere logic, nothing can be considered as rational – the agent can never fail to do something in a rational sense, solely in a factual sense, and then it should probably be called a mistake – not an act of irrationality. Consider the above example again, now however the agent is using a screwdriver, because she holds the belief which identifies this as the proper means of bringing around her desired end. On the surface this agent appears to be rational, because of the correspondence between her belief and her desire (getting the screw into the wall), however is this factual correspondence, at all, an expression of rationality? In accordance with Ken Binmore’s (2009: 6) modern approach to rational choice, one might also argue that the choice depends upon how one wants to drive the screw into the wall – in a sense, then, it is a matter of *instrumental preferences*. An instrumental preference is state dependent in the sense that if it looks like raining one may prefer an umbrella to a piece of chocolate – that is, at one occasion one might prefer A over B, while in other situations one might prefer B over A. On the other hand, some preferences are not state-dependent and influenced by external factors (Binmore, 2009: 6).

Moreover, one could question the very foundation of instrumental rationality, because it solely addresses the rationality of action in regard to a particular means-end relation. In juxtaposition, the Platonic account of rationality (and the Kantian as I will discuss at some length later in this text) took into account the rationality of ends as the very foundation of rationality. If one applies a pure instrumental rational perspective, it becomes impossible to say anything about *why* the agent is pursuing a particular end, and as such the end is exogenous, and must be presupposed, just as the motivation to act must follow directly from the means-end relation – thus it must also, to some extent at least, be presupposed (Korsgaard, 2008: 43). However, if action is ever to be distinguished from mere behavior, it must be viewed as deliberative, and as constituting – hence it must be based on a rational foundation which encompasses that agents holds belief which make them follow a particular end, just as they hold beliefs which guides their practice, and that they are free to connect the means-end relations anyway they want insofar as it is in accordance with their reflexive rational judgments. Additionally, as will be made clear later in this text, if rational necessity should bear any meaning at all as a normative foundation it cannot be based on causal beliefs expressing necessity merely as anticipation (Korsgaard, 2008: 42).

### **3.3.5 Immanuel Kant**

The reconstruction of Kant’s conceptualization of rationality put forward here, will be divided into three steps; step one: at this step it is important to recognize the significance of Kant’s understanding of autonomy as put forward in his “Critique of Pure Reason” because it generates the foundation of his practical philosophy on rationality, just as it moves the analytical perspective from the results of the action to the intention behind this – that is, agents cause something to happen, not actions; step two: if agency is viewed as constitutive of actions, then it is important to understand the foundation of agency – that is, the principles on which this constitution is founded; and step three: having clarified the significance and meaning of autonomous self-constitution, the necessary foundation of rationality has also been explained, thus Kant’s conception rationality can be clarified.

“Freedom, in its practical sense, is the independence of our will from coercion through impulses of sensibility. Our will is sensible insofar as it is affected pathologically (...) it is animal will (...) if it can be necessitated pathologically. The human will is certainly an *arbitrium sensitivum*, not, however, *brutum* but *liberum*, for sensible impulses do not necessitate its action, but there is in human beings a faculty of self-determination, independent of the necessitation through sensible impulses.” (Kant, 1781/1787/2007: B562,563/A534, 535; 464)

What is important to notice in this quotation is that, the act of the agent is unconditioned, that is, the decision to act plays a causal role, even though it is outside the natural causal relation. In that sense, the cause of the action is the acting agent’s decision to act, hence if the action does not produce the desired result it is not, as such, the fault of the actions – rather, it is the agent’s fault (Korsgaard, 2009: 83). An action is, then, a beginning – an initiation of a new causal series – however, the agent, as such, is not acting in a temporal/spatial vacuum (Allison, 1990: 26). Korsgaard (2009: 89) adds to this interpretation by explicating the inherent complexity of this autonomous condition, that is, autonomy is related to forward causation (the future in front of the agent), whereas natural causality is related to the past as cause and the present as effect. In that sense, *to be the cause* without *being caused* is the meaning of autonomous agency.

As such, one may question the existence of such an unconditioned act, however, taking into consideration Kant’s cosmological theory and his critique of Hume’s argument on causality. Hume’s argument can be reconstructed as follows: causal relations between external objects, as such, can solely be treated as mental habits following from memory, and actual sensory experience of these causal structures. The problem with this perspective is, according to Kant, that it presupposes a start of the causal series, because if one event is the effect of some previous event (the cause), this event must also have been caused by something (Kant, 1781/1787/2007: B561/A533, 463)<sup>22</sup>. What Kant is now arguing, is that the idea of transcendental freedom is just as impossible to experience as Hume claimed the idea of natural causality to be:

---

<sup>22</sup> The original statement goes: “The former is the connection of one state in the world of sense with a preceding state, upon which it follows according to a rule. As the causality of appearances depends on conditions of time; and as the preceding state, if it had always existed, could not have produced an effect which first arises in time, it follows that the causality of the cause of that which happens or arises must, according to the principle of the understanding, in turn itself require a cause.” (Kant, 1781/1787/2007: B561/A533; 463)

“...the faculty of beginning a state spontaneously. Its causality, therefore, does not depend, according to the law of nature, on another cause, by which it is determined in time. In this sense, freedom is a pure transcendental idea, which, firstly, contains nothing derived from experience; and secondly, the object of this idea cannot be given determinately in any experience, because there is a universal law of the very possibility of all experience, according to which everything that happens must have a cause, and according to which, therefore, the causality of the cause, which itself has happened or arisen, must also have a cause. In this manner, the whole field of experience, however far it may extend, has been changed into the sum total of mere nature. As, however, it is impossible in this way to arrive at an absolute totality of the conditions in causal relations, reason creates for itself the idea of spontaneity which can begin to act of itself, without an antecedent cause determining it to action, according to the law of causal connection. (Kant, 1781/1787/2007: B561/A533, 463)

This quotation states Kant’s idea of transcendental freedom in terms of the unconditioned event is closely related to the statement above on unconditioned actions, because unlike Hume’s perspective, the autonomous act is not caused by any external factors or internal factors such as inclinations or desires (at least not moral actions), or perhaps a more commonsense interpretation is that, human agents are provided with the freedom not to act on these desires. In that sense, when agents choose to act, the action, as such, does not *arise* from something, but it *follows* upon something (reason), as Allison argues (1990: 26). Kant’s perspective, then, assumes agency – action which springs from the agent – and because these actions springs from the autonomous agent they must, as such, be representations of the nature or character of the agent and, simultaneously, constitutive of these faculties (Allison, 1990: 28). In juxtaposition to, for example, Hume’s perspective, this perspective does not take into consideration the cause of the action, that is, Hume’s perspective emphasized the existence of sufficient reasons, both external and internal, as the causal origin of actions. Moreover, a theory which presupposes autonomous agency must, to some extent at least, also presuppose a responsibility which extent the consequences of one’s actions – a responsibility for one’s motives, and as such the Kantian subject is, necessarily, committed not just to her actions but also her motives (Korsgaard, 2008: 57; 2009: 87). Although, one should not interpret this, as Korsgaard (2008: 50) argues, in such a way that the agent necessarily is motivated to will the actions, just because she wills the end, rather both properties must be viewed as being founded on a deliberate choice.

To understand Kant's concept of will and its relation to autonomy, it is important, or at least it argued here that it is, to understand his project – what he sought to clarify or, at least explicate, was the metaphysical foundation of morality and rationality, thus its motivational foundation. Perhaps another, but closely related, interpretation is that he sought the foundation of judgment on which morality and rationality should (in a normative sense) be founded – that is, the foundation of justifiable reasons on which morality and rationality should be build (Korsgaard, 1996: 116). This seems to be the case at the end of the first part of his work “Grounding for a Metaphysics of Morals” when he argues:

“Since I have robbed the will of every impulse that could have arisen from the obedience to any law, there is nothing left over except the universal lawfulness of the action in general which alone is to serve the will as its principle, i.e., I ought never to conduct myself except so that I could also will that my maxim become a universal law. Here it is mere lawfulness in general (without grounding it on any law determining certain actions) that serves the will as its principle, and also must so serve it, if duty is not to be everywhere an empty delusion and a chimerical concept; common human reason, indeed, agrees perfectly with this in its practical judgment, and has the principle just cited always before its eyes.” (Kant, 1785/2002: 4:402; 18)

What Kant is arguing here, it seems, is that morality should be founded on the will of the agent, however, a will stripped of its subjective reasons, i.e. inclinations (subjective desires), that is, causes which properties are not such that they could be taken as objective – transcending the experience of the single agent. The agent should act, on *a priori* reason, with no motivation besides her obligation to the objective law. This is a distinct kind internalism. Nevertheless, this quotation also shows that, the agent is assumed, ideally at least, to be in possession of, or rather perhaps, to be capable of employing her free will, albeit it is, as such, not entirely free but governed by the subjective principles of the agent, and the objective imperatives – a distinction which will be primary focus in the next subsection as well. For now, the primary focus is Kant's concept of free will. This quotation also illuminates the implicit complexity of the Kantian concept of autonomy, because on the one hand, it is a property of the will, and on the other hand, it is the foundation of morality (Allison, 1990: 94). In regard to the former conception of autonomy, Kant defines it as a property of the will:

“Autonomy of the will is the property of the will through which it is a law to itself (independently of all properties of the objects of volition). The princi-

ple of autonomy is thus: ‘Not to choose otherwise than so that the maxims of one’s choice are at the same time comprehended with it in the same volition as universal law’.” (Kant, 1785/2002: 4: 441; 58)

That is, autonomy as the property of the will means to act in accordance with and from the principles governing the will which are, again, self-governing principles, thus the subjective principles which governs the actions are equal to the action (in so far as the agent acts in complete accordance with them), which again is equal to the will (the intention) and the goal of the action – which as such is nothing but the action in itself. This also seems to be argued in this quotation:

“The last must therefore abstract from every object to the extent that it has no *influence* on the will, hence practical reason (will) does not merely administer some other interest, but merely proves its own commanding authority as supreme legislation.” (Kant, 1785/2002: 4: 441; 59)

Moreover, the autonomous will is a *law to itself* which implies, in the lines of the “Critique of Pure Reason”, that it is causally undetermined by any external events, and as such solely causally connected to the will itself – that is, it is spontaneous. Kant argues that this should be understood as contrasted to, as already argues, causally connected actions – or what he calls the heteronomy of the will:

“If the will seeks that which should determine it *anywhere else* than in the suitability of its maxims for its own universal legislation, hence if it, insofar as it advances beyond itself, seeks the law in the constitution of any of its objects, then *heteronomy* always comes out of this. Then the will does not give itself the law but the object through its relation to the will gives the law to it. Through this relation, whether it rests now on inclination or on representations of reason, only hypothetical imperatives are possible: ‘I ought to do something *because I will something else*’. By contrast, the moral, hence categorical, imperative says: ‘I ought to act thus-and-so even if I did not will anything else’.” (Kant, 1785/2002: 4: 441; 58)

The heteronomy of the will, then, as a property of the will, is that which can be identified in the perspectives of Hobbes, Smith (to some extent) and Hume – that is, when the action has as its aim not the will in itself, but something other than this – when the end is exogenously determined and the action is directed at obtaining something besides itself and, as such, the motivation of the action is distinct from the action itself. The motiva-

tion *A* of action *X* is the obtainment of *B*, compared to the autonomous perspective in which, the motivation *A* is equal to the goal *A* of the action. Furthermore, the motivation and the goal is separated because in the heteronomous perspective, the agent acts on goals which are external, just as her motivation to act stems from external stimuli – hedonic experience. In other words, the will of the agent is not, in this perspective, autonomous, but determined by external events. In a modern perspective, inspired by this framework, developed by Edward Deci and Richard Ryan (2006: 6), the authors argue that heteronomy may also be associated with regulation from the outside which might, in some cases, be experienced as alienating by the agent, in the sense that external stimuli, such as contingent reward or punishment may out-crowd true internal will. As I will argue in chapter VIII, autonomy should not be, at least not according to Deci and Ryan (2000), be viewed as equivalent to social independence.

As Allison (1990: 97) argues, this does not imply causal independence or freedom from these external events of experience – autonomy, he argues, lies in the property of the will to self-determination, what above was referred to as spontaneity, that is, to act on reasons which are independent of or even contrary to this external stimuli.<sup>23</sup> By doing so, the agent is passing a tacit claim on the world, a claim on being respected and treated with dignity – for short, a claim founded on mutual trust (Korsgaard, 2009: 87), I return to this below. Furthermore, it also implies that the agent is not, as such, in complete control of the results of her actions, because by choosing to be the cause, the agent is also choosing to put herself within a causal chain which, as I argued above drawing on Korsgaard’s interpretation, implies that the agent must not only will the end as an abstract idea, but also will to actually bring the end around, and as such she is responsible for the actions, but it also implies that this commitment presupposes belief, a belief about the properties of the end, and the properties of the action, and to some extent also *faith* – a belief on the power of causation – that is, the agent must *believe* that it actually matters, that, for example, her moral actions actually are acknowledged as such, and

---

<sup>23</sup> Kant, then, at the beginning of the third chapter of “Groundwork to Metaphysics of Morals”, distinguishes between positive and negative freedom – the positive springs from the negative he argues. The negative concept of freedom, he argues, is that rational agents are equipped with reason, thus they have the ability to act independently of natural (external) causes. The positive concept, then, implies that the agent as an autonomous agent is governed by her self-constituted laws, however, laws which has the possibility of being made objective – universal – what he calls a particular kind of laws – laws which springs from the agent’s reflexive judgment (Kant, 1785/2002: 4: 447, 63).

that she is actually contributing to making a difference, thus it implies a belief on the efficacy of the action (Korsgaard, 2009: 88). I return to this below.

As I mentioned above, the autonomy of will also takes on a moral nature, because as it is argued in the quotation above, morality is based on the categorical imperative which implies that the motivation of moral actions is the action itself. This perspective on moral motivation is in juxtaposition to the perspectives of Hobbes who argued that the foundation of morality was: on the one hand the happiness of the agent; and on the other the primary aim of rationality, self-preservation; Hume argued that the foundation of morality was hedonic experience – feelings of solidarity – and Smith added to this moral feeling, virtues – actions which has the benefit of the other as its goal. Smith’s perspective is, to some extent at least, related to Kant’s, although; Kant would not have acknowledged the emotional foundation of morality. According to Kant, the foundation of morality – the constitutive reasons – cannot be one’s own happiness, because, like Plato did, he distinguishes between being happy and being good, and similar to Smith, he critically argues that happiness is related to prudence – that is, when the reflexive power of the rational faculty is directed at maximizing the happiness of the individual, and as such, applies its mental capacities to draw better calculations, without distinguishing between vices and virtues (Kant, 1785/2002: 4: 442; 59-60; 1788/2000: 26 [Theorem II]:)<sup>24</sup>. Kant argues in regard to emotions as the foundation of morality that, even though, they are preferable to happiness as the foundation, they could not generate a proper foundation because it is impossible, through introspection, to understand the other, just as emotions cannot be used as the foundation of judgment (Kant, 1785/2002: 4: 442; 60). Kant clarifies this argument in his later work, “Critique of Practical Reason”:

“The principle of private happiness, however much understanding and reason may be used in it, cannot contain any other determining principles for the will than those which belong to the *lower* desires; and either there are no higher desires at all, or pure reason must of itself alone be practical; that is, it must be able to determine the will by the mere form of the practical rule

---

<sup>24</sup> In the sections on Hume and Smith it was argued that self-esteem was derived from how other social agents relates to oneself. From a Kantian perspective, this relative conception of self-worth which is external in causality cannot be defended, insofar as this is a false sense of self-esteem. True self-esteem in a Kantian sense, then, comes from within and is not, as such, contingent upon external factors, but stems from a true conception of oneself which can only be displayed in volition – choices which springs truly from the agent’s own will (Deci and Ryan, 1995: 35).

without presupposing any feeling, and consequently without any idea of the pleasant or unpleasant, which is the matter of the desire, and which is always an empirical condition of the principles. Then only, when reason of itself determines the will (not as the servant of the inclination), it is really a higher desire to which that which is pathologically determined is subordinate, and is really, and even specifically, distinct from the latter, so that even the slightest admixture of the motives of the latter impairs its strength and superiority; just as in a mathematical demonstration the least empirical condition would degrade and destroy its force and value. Reason, with its practical law, determines the will immediately, not by means of an intervening feeling of pleasure or pain, not even of pleasure in the law itself, and it is only because it can, as pure reason, be practical, that it is possible for it to be legislative.” (Kant, 1788/2004: 23-24 [Theorem I, Remark I])

Morality is, then, founded on rationality – that is, the reflexive capabilities of the autonomous agent. The agent’s principles (motivation) are not influenced by any material or external incentives, but entirely based on the will of the agent. The actions of the agent must follow her will, which springs from her principles which in turn must be governed by the principles of universality – that is, the categorical imperative. The relation between the will and the action is a matter of internal motivation, while the relation between the will and the categorical imperative is a matter of pragmatic necessity (duty), and the relation between the principles and the categorical imperative is a matter of rational autonomy (or combined it generates what Kant calls the practical sense) (Kant, 1788/2004: 37).

### ***3.3.5.1 Self-Governing Agency***

Above rationality was tentatively defined as the reflexive power of the agent – that is, as the capability of the agent to generate through reflexivity the principles of her will, and to govern her actions. In the section above, it was also identified that Kant distinguishes between subjective maxims and objective laws; and between categorical imperatives and hypothetical imperatives. In his “Groundwork for a Metaphysics of Morals” Kant provides two different definitions of the distinction between subjective maxims and objective laws:

The first definition:

“A maxim is the subjective principle of the volition; the objective principle (i.e., that which would serve all rational beings also subjectively as a practi-

cal principle if reason had full control over the faculty of desire) is the practical law.” (Kant, 1785/2002: 4: 401n; 16n)

The second definition:

“A *maxim* is the subjective principle for action, and must be distinguished from the *objective principle*, namely the practical law. The former contains the practical rule that reason determines in accord with the conditions of the subject (often its ignorance or also its inclinations), and is thus the principle in accordance with which the subject *acts*; but the law is the objective principle, valid for every rational being, and the principle in accordance with which it *ought to act*, i.e., an imperative.” (Kant, 1785/2002: 4:421n; 37n)

According to the two definitions above, a maxim is subjective in nature, which implies that it is not, as such, universalizable; rather, it is a practical rule which governs the actions of the agent in accordance with her desires and condition – that is, they are private norms signifying a generalized motivation and governing of actions, applied within relevant and similar contexts (Allison, 1990: 87). That is, “...*principles are general outlines within which details find a place...*” (Nozick: 1993: 38). In that sense, a maxim is a first-order condition of actions which *determines* how the agent acts. These maxims are, as such, not handed to the agent by either some divine force or the practical law, rather as Kant argues, they are self-imposed rules (Kant, 1785/2002: 4: 438; 56) created through practical assessment, and as such, the agent should not follow them blindly, but take into consideration the context – that is, “in contexts of type A, do actions of type D”. Additionally, this also seems to imply, at least some, level of self-consciousness – that is, the agent who chooses the principles of her actions must also, at least to some extent, be aware of these principles governing her actions. Robert Nozick (1993) argues on this matter:

“The Kantian tradition tends to hold that principles function to guide the deliberation and action of self-conscious creatures, hence principles have a theoretical and practical function. We are creatures who do not act automatically – without any guidance.” (Nozick, 1993: 38)

Imperatives, on the other hand, are objective laws governing the normative foundations of how the agent should act. As such, an imperative is a second-order principle of action. In other words, while the first order principle commands “do X-type actions”, the second-order principle governs the first-order principles, and as such it is solely reachable through reflexivity (Allison, 1990: 90). It is the rational determination of appropriate

actions which then lend itself to the will as motivation and possible appropriate actions through the working of private norms. Appropriateness, may in this context, be understood as either internal, autonomous, and universal appropriateness, thus signifying that the imperative is categorical, or be limited to the condition, the inclination, or for short the context, which thus implies that the imperative is hypothetical (Allison, 1990: 89).

### ***3.3.5.2 Kantian Rationality***

To understand Kant's idea of rationality it is necessary to distinguish between two different types rationality, non-moral rationality associated with means-end relations, and social rationality associated with will-end relations and morality. Additionally, it is necessary to distinguish between three different forms of rationality: weak rationality associated with instrumental rationality; semi-strong rationality associated with instrumental rationality founded on a common goal, happiness, which then generates harmony; and a strong form of rationality in which the ends are set by the agent, thus rationality is not associated with proliferation, but with goal setting. The latter form of rationality is the only type which can be considered as the proper foundation of social rationality. To get a better understanding of these important distinctions, the aim of this section is to discuss each individual form and type of rationality identified by Kant.

#### ***3.3.5.2.1 The Weak form of Rationality***

The weak form of rationality is only briefly identified by Kant as a form of rationality which is entirely based on the pursuit of empirical desires without any common goal, that is, rationality here refers solely to the process of identifying the actions which are, viewed by the agent, as optimal in regard to bringing around some desired future ends (Korsgaard, 2008: 47; Kant 32). These goals which the agent pursues, are not set by the will of the agent, but are exclusively the creation of nature – empirical and material in nature – which presents itself to the agent as inclinations – that is, as desires. The role of reason, in regard to this form, is solely to guide the choice of actions, what Kant (Kant, 1785/2002: 4:416; 32) calls imperatives of skill. Skill can be defined as the developed (or learned) ability to carry out predetermined goals with a minimum use of resources (i.e. time and energy), which entails that reason, here, is associated with the

choice of action in relation to some end, and the procedural nature of the action. In other words, reason is to be considered as technical in nature, and the actions identified by reason as proper in relation to some goal are to be considered as conditionally necessary. That is, another agent may not necessarily identify the end as something worth of pursuing, just as she may choose different means to pursue it based on her practical reason. As such, the domain of practical reason associated with the weak form of rationality is the actual or factual desires and resources of the agent. Irrationality, then, takes on a Humean dimension as actions which are based on wrong beliefs about the relation between means and ends, or the actions are ill-fitted to the actual situation of the agent, I return to this below.

This weak form of rationality is a generic form of rationality which is the point of departure for most of the perspectives on rationality. The next step is to develop a particular form of rationality in which limitations of some sort are introduced most often in form of a general principle of rationality. One author, however, deviates from this way of reasoning, Montaigne. According to Montaigne, rationality is malleable to the desires of the mind, and the only limitations imposed on desires are the ones acquired through life. Of course, Montaigne, like the Stoics before him, believed in a concept of rationality which was not, entirely that is, directed by proliferation and practical concerns, but rather aimed at identifying a rational way of life. This particular understanding of rationality was, to some extent at least, abandoned by later authors on the subject, perhaps inspired by Hobbes who identified rationality as a principle of action and social conduct. Kant's understanding of rationality is, to some extent, broader in scope because it takes into consideration not only the politics of one's conduct, but also that rationality refers to which goals should be followed, not in terms of *true* goals, as Plato, nor *immoral* goals, as Hume and Smith, but rather in terms of authentic goals which springs from the pure will of the agent, see the section on the autonomy of will above. According to Kant, these authors based their understanding of rationality on a kind of predetermined, or natural, understanding of ends, and considered, as a result, solely limitations of rationality as those which were necessary due to some empirical reality which the agents necessarily were a part of. In order to escape this, Kant proposed a kind of rationality which was entirely based on autonomy, as I have already argued above.

However, before I discuss this, it is necessary to discuss a mediating form of rationality which is close to the form of rationality presented by Plato and Bentham.

### 3.3.5.2.2 The Semi-Strong form of Rationality

The semi-strong form of rationality identified by Kant refers to a form of rationality in which the ends are still empirical in nature, that is, the end and the will of agent is separated analytically. However, in juxtaposition to the weak form, this form is centered upon a common and harmonizing goal, happiness (Korsgaard, 1996: 111; 118). Kant argues:

“There is *one* end, however, that one can presuppose as actual for all rational beings (insofar as imperatives apply to them, namely as dependent beings) and thus one aim that they not merely *can* have, but of which one can safely presuppose that without exception they *do have* it in accordance with a natural necessity, and that is the aim at *happiness*.” (Kant, 1785/2002: 4:415; 32)

Plato argued that well-being is a harmonious goal which induces a kind of consistency in action and imposed a third-order principle of desire on the choices of the agent – that is, a choice is rational insofar as it is consistent with the well-being of the agent. Hobbes argued, in a similar way, that the primary principle of rationality is maximization of power governed by the rational principle of self-preservation. By arguing that a common goal is happiness, Kant is also arguing that the role of reason is not solely to be identified with instrumental rationality, because the goal must also, in some way, contribute not only to the present, but also the future happiness of the agent. Reason is, here, to be considered as *prudence*, that is, the ability of the agent to, through the foresight of the rational faculty, reason about what is virtuous, appropriate, and profitable. Kant argues, however, that prudence and happiness cannot be understood as an objective end, because the idea of happiness is ambiguous:

“Yet it is a misfortune that the concept of happiness is such an indeterminate concept that although every human being wishes to attain it, he can never say, determinately and in a way that is harmonious with himself, what he really wishes and wills. The cause of this is that all the elements that belong to the concept of happiness are altogether empirical, i.e., have to be gotten from experience, while for the idea of happiness an absolute whole, a max-

imum of welfare, is required, in my present and in every future condition. Now it is impossible for the most insightful, and at the same time most resourceful, yet finite being to make a determinate concept of what he really wills here. If he wills wealth, how much worry, envy, and harassment will he not bring down on his shoulders?" (Kant, 1785/2002: 4: 418; 34)

Happiness is a condition, a state of mind, and as such it is partly will and partly wish. That is, in juxtaposition to the strong form of rationality, the foundation here is not entirely the will of the agent, but also wishful-thinking and imagination – what the agent hopes. The finite agent, nevertheless, cannot properly determine what will contribute to her well-being, because the consequences of the means belong entirely to empirical causes and, as such, she cannot thoroughly determine all the possible consequences, just as what is expected to bring happiness must be founded on experience which, as such, is limited to the experience – her lived life. In that sense, the consequences are beyond the control of the agent, just as the choice of means is based on an uncertain epistemological foundation. Thus, the imperative of reason cannot be *law*, because that would imply an objective/universal end, just as it would imply objective means with certain results. Additionally, the foundation of happiness is not entirely rationality, but also wishful-thinking, hence prudence can solely *guide* the actions of the agent. In other words, the role of reason is, here, to give counsel to the rational agent – that is, to give advises to the rational agent in regard to which ends she should pursue, and, in an instrumental sense, which actions she should choose (Kant, 1785/2002: 4: 418-419; 35). An end, in this perspective, is considered to be rational as long as it is consistent with the expectations of agent in regard to what she expects and believes contribute to her well-being.

#### 3.3.5.2.3 The Strong form of Rationality

Kant, like Plato, also identifies a strong form of rationality – a form which is based on the constitution of authentic goal, albeit Plato would also add true goals. This latter criterion on *trueness* would, however, make no sense to Kant because to him the metaphysical foundation of this strong form of rationality is not empirical, but a priori and, as such, the foundation of will is autonomy not heteronomy; see the previous section (see also below). According to Kant, an agent should only pursue her authentic ends, that is, the end which is in accordance with her will, or more precisely – an end

which is constituted by her own will. Furthermore, she should only pursue an end which can be regarded as an objective or universal end – an end which she perceives all other rational agents would perceive as a proper end. Contrary to the two other forms of rationality, I discussed above, this form of rationality is based on the categorical imperative – that is, an unconditional law which governs all rational agents in all situations, thus it is not as the technical or prudential imperative hypothetical, conditional, in nature. Kant argues:

“...that the categorical imperative alone can be stated as a practical law, while the others collectively are, to be sure, *principles* of the will, but cannot be called ‘laws’; for what it is necessary to do for the attainment of a discretionary aim can be considered in itself to be contingent, and we can always be rid of the precept if we give up the aim; whereas the unconditioned command leaves the will no free discretion in regard to the opposite, hence it alone carries with it that necessity which we demand for a law.” (Kant, 1785/2002: 4:420; 36 italics in original)

In addition, the categorical imperative should be understood as separated from a particular end, and solely related to the will of rational agents, thus a categorical imperative do not, as such, disappear if a certain end is removed. In other words, it is *not* particularistic in nature (Korsgaard, 2008:121; 2009:73) Kant goes on discussing the nature of this certain form of rationality as distinguished from the other two forms:

“That which, by contrast, is derived only from what is proper to the particular natural predisposition of humanity, or from certain feelings and propensities, or indeed, if possible, from a particular direction of human reason, and would not have to be valid necessarily for the will of every rational being—that can, to be sure, be a maxim for us, but cannot yield any law; it can yield a subjective principle, in accordance with which we may have a propensity and inclination, but not an objective one, in accordance with which we would be *assigned* to act, even if it were to go directly contrary to all our propensities, inclinations, and natural adaptations; it even proves all the more the sublimity and inner dignity of the command in a duty, the less subjective causes are for it and the more they are against it, without on this account the least weakening the necessitation through the law or taking anything away from its validity.” (Kant, 1785/2002:4:425; 43)

As Kant seems to argue in this quotation, the foundation of this form of rationality is the autonomous will of the agent, and the imperative which govern such an autonomous

will must be of such a kind that its legitimacy is not limited to certain situations, just as it must be legitimate even in situations in which it seems to be contrary to the *nature* of the agent – her past actions, her desires/inclinations, and sentiments. In that sense, the maxims based on the autonomous will of the agent, should not just be considered as *ad hoc* rules or particular rules guiding particular means-object relations. Rather, they should be viewed as binding to the agent – self-binding that is – demanding as I argued above, commitment on behalf of the agent; a commitment which also seems to follow from the autonomous nature of the will. In other words, the autonomous condition of humanity which once freed them from the chains of nature now confines them to self-governance and a continuous demand to act (Korsgaard, 2009: 87). Contrary to Plato’s idea of the will which was centered upon an emotional foundation (i.e. indignation) and the foundation of the higher desires (i.e. a desire for justice), Kant’s idea of the will is more related to self-command and self-control. This kind of self-binding should be understood as the bounds of reasons, that is, the boundaries of the will are centered upon reason, and as such not by any particular means-*object* relation (Kant, 1785/2002: 4:427; 44). Kant argues:

“Here, however, we are talking about objectively practical laws, hence about the relation of a will to itself insofar as it determines itself merely through reason, such that everything that has reference to the empirical falls away of itself; because if *reason for itself alone* determines conduct (the possibility of which we will investigate right now), it must necessarily do this *a priori*. The will is thought as a faculty of determining itself to action *in accord with the representation of certain laws.*” (Kant, 1785/2002:4:427; 45)

As such, the will is only bound by reason, and as such it is autonomous with regard to empirical and external incentives. The boundaries of reason are those which govern all rational agents, and should not be understood as external and coercive in a structuralistic sense, but rather as a kind of self-binding which refers to the limitations stated in the categorical imperative, that an action should be fit to serve as a universal law governing all rational agents. The will refers as such solely to itself, thus the role ascribed to the Kantian agent is threefold: she is *the author* of her own ends, she is *the legislator* of her own laws (i.e. imposes sanctions), and she is *the subject* to her own laws (Kant, 1785/2002: 49). In that sense, the will is a self-referential system based on reflexive

rationality. The table below presents the three different forms of rationality and their respective differences.

**Table 1: Summary of the discussion on the basic features of the three kinds of Kantian rationality.**

	<b>Weak Rationality</b>	<b>Semi-strong Rationality</b>	<b>Strong Rationality</b>
<b>Reason</b>	Technical	Prudence	Volition
<b>Imperative</b>	<i>Hypothetical</i>	Hypothetical	Categorical
<b>Role of Reason</b>	Rules of Skill	Counsel of prudence	Commands of Morality
<b>Metaphysics of Reason</b>	Conditional practical necessity [actuality]	Conditional pragmatic necessity [possibility through imagination – wishful-thinking]	Unconditional pragmatic necessity [possibility through reason]
<b>Metaphysics of Will</b>	Empirical	Empirical	A priori
<b>Actions</b>	Means to an end	Means to an end	End in itself
<b>Foundation</b>	Desire	Happiness	A priori will
<b>Foundation of Will</b>	<i>Heteronomy</i>	Heteronomy	Autonomy
<b>Principle of Action</b>	First order	Second order (based on happiness)	Third order (self-constitution)
<b>Value(s)</b>	Pain and pleasure	Happiness, harmony, sentiments	Dignity, respect, authenticity, integrity
<b>Foundation of justification</b>	Consequences	Consequences	Generative principle/motive
<b>Locus of Worth</b>	Extrinsic	Extrinsic	Intrinsic
<b>Area of Legitimacy</b>	Means-end relations	Means-Happiness/harmony relation	Will-end relation

### **3.3.5.3 Social Rationality within the Bounds of mere Reason**

To some extent, this description of the will seems to point to a particular role of agency as the foundation of society, and that individuals are not bound to each other by anything but the objective laws which, as such, solely exist as a priori laws of reason. Even though, this conception is not entirely wrong, it is necessary to take into consideration the social relational dimension of Kantian agency and its relation to social rationality. Kant's understanding of social rationality is very different from the perspectives of Hobbes, Hume and, to some extent, Smith, because he does not take into consideration that the foundation is empirical, as Hume and Smith, and neither does he center it upon consequentialistic terms. The rational foundation of social rationality is the strong form of rationality, which implies that the metaphysical foundation of morality is the autonomous will of the agent. That is, a will which is, as such, exclusively limited by its self-constituted laws of reason, hence the moral action springs from the will of the agent

as a pure motive which is an end in itself. The agent, then, is moral as a result of her own will, not because she fears the consequences or because her moral conduct makes her feel good. This also entails that social rationality is authentic, insofar as it is a result of the autonomous will. It might, of course, be questioned if it is possible to base social rationality on reason and expect it not to be *perverted* by empirical incentives, just as it might be questioned if reason can ever be autonomous, at least in regard to social norms. Kant, of course, was aware of this and describes his philosophy of social rationality as an ideal type not necessarily found in the real world.

However, what is important to recognize is that Kant by basing his theory on agency not only impregnates the agents with a social capability beyond desires, social norms, disguised egoism, and religion, he also turned the metaphysical foundation of morality from naturalistic to anthropocentric in origin. That is, just as he argued that the foundation of cognition was not experience, in the Humean conception, but rather that the order perceived in nature was the result of the faculties of the agent, his moral philosophy also states that the laws of morality, as such, are not existing external to the agents – the foundation of good and evil is not separately existing and external laws, either based on religion, social convention or moral sentiments (Kant, 1793/1794/2009: 31). On the one hand, Kant is freeing the agents from the empirical chains forged by naturalism and emotivism, and on the other hand, it seems that he is confining them to their own moral judgment and, as such, commits them on themselves and the self-imposed bounds of reason.

The next step Kant made after he situated morality within the subject was to clarify the psychological foundation of this particular kind of social rationality. The rational agent, Kant argues, is by his self-command a member of a kingdom ends (Kant, 1785/2002:4:432; 51). Kant adds on this:

“For rational beings all stand under the *law* that every one of them ought to treat itself and all others *never merely as means*, but always *at the same time as end in itself*. From this, however, arises a systematic combination of rational beings through communal objective laws, i.e., a realm that, because these laws have as their aim the reference of these beings to one another as ends and means, can be called a ‘realm of ends’ (obviously only an ideal).” (Kant, 1785/2002:4:433; 51)

The origin of this kingdom of ends is social rationality, because the socially rational agent takes into consideration the other social agents, as equal autonomous wills, that is, when she designs her ends she is continuously asking herself if she will this end should serve as a general end, or her law as a general law. In other words, would the agent like if others acted similar to her. Furthermore, this entails that the socially rational agent must acknowledge that other social agents are targets of her ends, just as they may serve as means in regard to their mere implementation in practice – that is, while the actions refers to the will, and the will to reason, the reason is not, as such, self-referential but takes into consideration the objective law which binds all rational agents as an act of transcendent rational imagination. As such, the other social agent is included as a part of the autonomous will of the agent through her reasoning. Kant argues further on this matter:

“But a rational being belongs as a *member* to the realm of ends if in this realm it gives universal law but is also itself subject to these laws. It belongs to it *as supreme head*, if as giving law it is subject to no will of another. The rational being must always consider itself as giving law in a realm of ends possible through freedom of the will, whether as member or as supreme head. It can assert the place of the latter, however, not merely through the maxim of its will, but only when it is a fully independent being, without need and without limitation of faculties that are adequate to that will.”  
(Kant, 1785/2002: 4:433-34; 51-52)

What Kant is describing here, it seems, is a kind of *elusive* systemic rationality which binds the social agents together while still allowing them autonomy of will – a kind of morality based on the idea of objective government through self-government, that is, governance without government (understood as an organ which has the privilege of binding its subjects to a particular reason) – a kind of governance through reason. As such, the basic rhetoric figure of social relations may be captured as a *subject-subject* relation, contrary to a *subject-object* relation – that is, a shift from a relational type based on the rational calculation of how the other contributes to the agent’s egoistic goal, to a relational type which identifies the other agent as an equal rational agent who is also an autonomous author of the objective rules, thus the agent must respect her as such. This respect is simultaneously the foundation of the social relation based on social rationality, and the limit of the will of the individual agent, because one agent should

not impose limitations on the autonomous will of another agent, hence respect becomes an innate and inescapable foundational right that the agents owe to each other as members of the kingdom of ends. Kant argues further:

“Reason thus refers every maxim of the will as universally legislative to every other will and also to every action toward itself, and this not for the sake of any other practical motive or future advantage, but from the idea of the *dignity* of a rational being that obeys no law except that which at the same time it gives itself.” (Kant, 1785/2002:4:434; 52)

The dignity of the autonomous agent, then, becomes the foundation and the limit of social rationality. Furthermore, considerations on dignity is not just directed towards other agents, but also inwards because the agent should also respect her own dignity as an autonomous will, that is, she should commit herself on her own life – a kind of commitment which is also assumed in regard to the act of self-government. In that sense, the worth of the action is intrinsic in nature (Kant, 1785/2002: 52-3). That is, in a double sense of the word, because the social rational act is to be considered as an end in itself – an act on pure internal motivation, the motivating element flowing this action is the value of the action itself to the individual. Kant, contrast this to the properties of empirical motives which are extrinsic in nature – that is, ends which are not pursued because they represent a value in themselves, but acquires their value in comparison with other ends, i.e. desires which represent a market price, or intellectual powers which represent an affective price (Kant, 1785/2002: 52-3). This way of distinguishing between intrinsic and extrinsic in regard to motive and motivation can be viewed as an early account of the modern distinction between the economic approach to motive and motivation, and the social psychological perspective associated with the critical perspectives I will discuss in chapter IV.

**Table 2: The motivational gap.**

	<b>Motive</b>	<b>Motivation (psychological value)</b>
<b>Intrinsic (Autonomy)</b>	The motive originates from the will itself.	The psychological value originates from the act of will in itself, i.e. the job itself.
<b>Extrinsic (Heteronomy)</b>	The motive originates from an empirical desire which is separated from the will.	The psychological value originates from something associated with the end, i.e. money or promotion.

Kant concludes on this matter:

“...for the worth of these principles does not consist in effects that arise from them, in the advantage and utility that they obtain, but rather in the dispositions, i.e., the maxims of the will, which in this way are ready to reveal themselves in actions, even if they are not favored with success. These actions also need no recommendation from any subjective disposition or taste, regarding them with immediate favor and satisfaction, and no immediate propensity or feeling for it: they exhibit the will that carries them out as an object of an immediate respect, for which nothing but reason is required in order to *impose* them on the will, not to *cajole* them from it *by flattery*, which latter would, in any event, be a contradiction in the case of duties. This estimation thus makes the worth of such a way of thinking to be recognized as dignity, and sets it infinitely far above all price, with which it cannot at all be brought into computation or comparison without, as it were, mistaking and assailing its holiness.” (Kant, 1785/2002: 4:435; 53)

As such, the worth of the moral act may be understood as an inner worth because it does not represent any value except that which is associated with the act itself. In that sense, morality should not be *praised* as something exceptional, rather it should be assumed as important part of rational human nature (Kant, 1785/2002: 4:398; 14) – it should not spring from demand, force or disguised selfishness, but represent the *pure* good will of the social agent. Of course, it might be argued that Kant’s moral psychological foundation is rather *cold* because it is centered upon reason and not emotions as, for example, the perspectives of Hume, Jean-Jacques Rousseau, and Smith. Nevertheless, one may also understand the Kantian approach to social rationality as an *update* of the platonic understanding, a perspective build on true morality, not however true in a scientific sense (true/false), but authentic and as a display of integrity – that is, as a true act of will (Korsgaard, 2008: 113). In the table below I have identified the main differences between non-rationality and social rationality.

**Table 3: On the distinction between moral and non-moral domains of actions.**

	<b>The Domain of Moral Action</b>	<b>The Domain of Non-moral Actions</b>
<b>Price of Action</b>	Inner worth (determined internally)	Relative worth; Market price/ affective price (determined externally)
<b>Locus of Worth</b>	Intrinsic	Extrinsic
<b>Psychological values</b>	Respect, authenticity, integrity	Pain and pleasure, happiness,

		harmony
<b>Origin of motivation</b>	A priori will (reason)[independent]	Empirical incentives [dependent]
<b>Motivation and interest</b>	“Take an interest”	“Acting <i>from</i> interest”
<b>Foundation of Judgment</b>	Reflexive rationality	Desire, inclination, emotions, happiness, human nature in general
<b>Role of the Individual</b>	Subject of laws, legislator of laws, author of laws – self-constitution	None – constituted by nature
<b>Area of legitimacy</b>	Social Relations (Kingdom of Ends), subject-subject relations in general	E.g. market relations, subject-object relations in general
<b>Principle of legitimacy</b>	Objective, universal laws	Subjective laws
<b>Type of rationality</b>	Social Rationality	Instrumental Rationality, semi-strong rationality
<b>Target of Ends</b>	The acting individual and other rational individuals	The acting individual
<b>The Role of Reason</b>	Commands of morality	Gives counsel on prudence, technical guide
<b>Other individuals</b>	Ends in themselves	Means to an end (worldly prudence)
<b>Actions in general</b>	Ends in themselves	Means to an end (private prudence)
<b>Imperative (commands of reason)</b>	Categorical	Hypothetical
<b>The Principle of Will</b>	Autonomy	Heteronomy
<b>Temporal Direction</b>	Present	Future
<b>The Foundation of Practical Choice</b>	Principles/maxims subject to the categorical imperative	Causality between means and end

### ***3.3.5.4 On the Powers and Limits of the Rational Will***

Above I analyzed the foundation of rational motivation within the philosophy of Kant. In this section, I turn to the practical implications of rational motivation and in particular the implications which follows from the fact that even Kantians are humans (perhaps all too human). In other words, the aim of this section is to discuss the limits of rational motivation in terms of the causation of will. Studies of the will is no longer only reserved for philosophers, but has recently become an area of analysis in the social sciences, i.e. in experimental neuroeconomics and theories on the cognitive psychology of human motivation. Hence, the aim of this section is, then, to apply Kant’s philosophy as a point of departure for discussing the limits of the rational will, by drawing on theories from, for example, neuropsychology. In the first part of this section, I will discuss the cognitive limits of the will by focusing on the causality of *willing*. In the second section, I will focus on the limits of the will in regard to overcoming temptations.

### 3.3.5.4.1 On the Cognitive and Practical Limits of the Causal Powers of Will

As I argued above, the point of departure for Kant was the autonomous and self-commanding social agent. Two of the main themes in Kant's practical philosophy seem to be: firstly; to free the will from its chains of naturalistic philosophy; and secondly, the practical mastery of this autonomous condition – how the agents should manage this autonomous will. However, Kant does not, at least explicitly, discuss the limitations of the will – that is, the fact that sometimes the power of will is not enough; sometimes it may even be self-defeating to try to bring a certain state around (Elster, 1983). The cognitive limits of the causal powers of the will refer to the domain of the individual (inward causality), that is, when the agent try, through mere will alone, to make something happen within herself, i.e. to bring forth within herself a certain state. The practical limits of the causal powers of the will, on the other hand, refers to the outward causality of trying to bring about a certain external state through the act of mere will alone. A further refinement of this distinction would be to add Kant's distinction between the categorical and the hypothetical imperative, because insofar as we are concerned with matters related to natural causality, i.e. bringing something about externally which is caused by something outside the agent, the imperative which should guide the process of deliberation is the hypothetical imperative, opposed to matters related to phenomena which originates from the agent herself, in this case the imperative guiding the agent's process of deliberation should be the categorical imperative. In that sense, it is related to the metaphysical foundation of the will – autonomy and heteronomy. Drawing on the discussion in section 3.5 the following table can be made:

	<b>Autonomy (being the cause)</b>	<b>Heteronomy (being caused)</b>
<b>Inward causality</b>	Self-constitution through agency – self-commitment.	Efficacy in achieving something by internal means.
<b>Outward causality</b>	Having <i>faith</i> in one's ability to bring about an external change.	Efficacy in achieving something by external means.

As such, this is not only a discussion of the causal relation between means and ends, because sometimes the agent may indeed choose the right means to a particular end but

still the result does not occur, because the very fact that the agent tried to bring the state around made her fail (Elster, 1983: 46). For example, Ego cannot make Alter fall in love with him, in fact if he tried to make her do that, she would probably leave him, and neither can he make himself fall in love with Alter, just as it is impossible for the agent to make herself believe something at will, because that would require that she *forgot* that she herself is the author of this new belief (Elster, 1983: 51, 53, 57-58, 62).

Some results, Elster (1983: 43, 44) argues, are entirely by-products of other results and as such cannot be brought around directly by the will of the agent, albeit this does not imply that agents do not sometimes will something which cannot be willed. This is, in less philosophical terms, a matter of causation – bringing something around – and of wanting to cause a particular effect, which as such is not only beyond the power of the agent, it seems also irrational to believe that it would be possible, because the cause is not, as such, directly linked to the effect which implies the existence of logical inconsistencies in the practical reasoning of the agent – although this does not imply that agent cannot anticipate what might be the consequences (Elster, 1983: 55). Elster argues on this matter:

“...they are states that cannot be brought about intelligently and intentionally. They may (...) be brought about knowingly and intelligently, if the agent knows that as a result of his action the effect will come about in a certain way. They may also come about intentionally and non-intelligently, if the agent achieves by fluke what he set out to bring about.” (Elster, 1983: 56)

For example, the employee who wants to impress his manager might find it problematic to reach a plausible strategy on how to succeed, because nothing impresses less than acts designed to impress (Elster, 1983: 66). Some authors on modern management, seems to believe that escaping the chains of the past to embrace the present implies leaving aside this mental presence of the past. However, this seems, at least, irrational to believe because the past, as such, is solely living as memory, thus the presence is nothing but a mental presence which is not reduced by trying to reduce it, that is, by bringing it present in the mind. In that sense, the agent cannot make herself forget something by willing to forget it – it seems to be a self-defeating act.

Consider another example – “pull yourself together”, the paradox here is not only a matter of logic but also of inward causation. That is, how can the agent be the cause of

her own effect – like Munchhausen who allegedly could pull himself and his horse from the swamp just by pulling his own hair? Agency presupposes the autonomous choice, and by choosing, the agent also constitutes herself as an autonomous agent whose identity is not caused by something outside her, but, as such, is her own creation – as a practical identity (Korsgaard, 2009: 129). Thus, self-constitution requires self-consciousness, just as it requires self-commitment and a belief on one's power of causality. Moreover, it also presupposes identification, that is, the agent chooses her actions in accordance with the unity of her will which implies that the process of deliberation, as such, is not only about identifying the optimal means-end relation, but also about identifying actions which are in accordance with, and contribute to this unity of the will (Korsgaard, 2009: 127). In other words, the process of deliberation is about identifying one's own way and choosing between different incentives each offering one to act in a specific way. The Kantian conception of the will is fixed by the principles that the rational agent applies in her process of deliberation and as such it is bound by these by commitment to act in a particular way. Now, the question seems to be, can the agent actually bind herself, that is, is it always feasible for her to command herself, is she always conscious of her own choice. Here it seems the discussion is entering the domain of the power of will a matter I return to below.

These examples seems to point to a will which is not only limited in regard to bringing something forth in another social agent, but it also limited in regard to bringing something forth within the agent himself. Thus, the will has cognitive limits in relation to reasoning about what is within its power. This is of course also related to the practical limitations of the agent in relation to her space of possible actions. Indeed, the will may entirely be limited by mere reason which implies that the agent is completely in control of her will-action relation, as such, she can carry out the intentions of her will. In other words, the will guided by reason deduces the possible actions from its own limitations brought upon it by reason, and not limits imposed on it by social structures as argued by structural functionalists. Additionally, insofar as Kant exclusively takes into consideration the pure autonomous will he is not limited by practical boundaries, and as a consequence focus entirely on the pure possibilities of the will, that is, what is within its power – a power determined by reason, and the mental faculties of the agent. Consequentialists, on the other hand, do not, as such, address possibilities, but rather opportunities,

understood in a double sense as; on the one hand, what is to be considered as an opportunity, that is the practical limits of actions determined by the external/empirical structures of actions; and on the other hand, as actions identified by the agent as opportune, that is as having desirable consequences (Elster, 2007: 165). Another implication of this distinction would be to consider the propositional attitudes implied in the Humean and Kantian sense as a distinction between command and intention – the Kantian subject can solely command herself, which implies that she would have carried out a particular action in the absence of an internal will, and as such the action is, inherently, intentional; while the Humean subject would not take on an action in the absence of some empirical incentive representing itself to the agent as an internal desire (Elster, 1983: 61). Sometimes these commands, either directed towards oneself or others, are illogical, in the sense that was explained in the text above, because they induce an illogical inconsistency in the relations intention-effect or cause-effect. For example, it is a pragmatic contradiction to command oneself to be spontaneous, just as it is to command the other agent to be more independent and not constantly follow the commands of other people.

First of all, as it is often the case, excessiveness leads to irrationality and a theory entirely centered upon rationality may be irrational, insofar as it seems to be centered upon an irrational conception of will-end reasoning. I return to this conception of irrationality in the section on Jon Elster and Donald Davidson. As it was also the case with the Platonic approach to rationality, the agent cannot be forced to accept, insofar as this acceptance has to spring from the free will of the social agent. In accordance with the model on the causal production of motives developed in the section on Smith, the Kantian conception of social rationality belongs to the social-motivated part of social rationality. However, the moral nature is not solely a result of its motivation, but also the claim which accompanies it, it must be willed to be elevated to a general law. As such, a moral norm is unconditional and accompanied by a normative belief on the quality of the value. The agent accepting a set of values implied in a compensation policy must, in other words, will that these come to bind her and others, just as these principles must be in accordance with her private values. Although, the approach developed by Kant can be described as a *cold* cognitive approach to rationality, it signifies that rationality is not the opposite of morality, but rather the foundation. The agent who accepts and binds herself to the principles produced by her will do so not because she has to, rather she

does it because she wants to. The claim produced by the social agent is solely a tacit claim on respect, and as such it is unconditional. The rational agent must, in other words, will for others what she wants for herself. The agent is able to transcend the immediate present in order to do what is right – and not just due to concerns about esteem and self-esteem. From what has been argued here, it seems that in order to understand Folger's conception of deonance in an intellectual historic perspective, one cannot separate it from its rational foundation, insofar as both Plato and Kant viewed reason and rationality as constitutive of morality. This, of course, presupposes a stronger kind of rationality than the one often invoked in modern literature, because if rationality is to be viewed as constitutive of morality, one needs to address a type of rationality which pushes past mere instrumental rationality, insofar as it demands that the social agents are equipped with an ability to reason about goals.

### **3.3.6 Jeremy Bentham and Economic Rationality**

Hume, and later J. S. Mill (1882/2009: 1151-1152), argued that the aim of rational choice, and the role of practical reason, was not to influence the process of goal formation, but to take these as given (exogenous), and reason about how these goals should be pursued by the rational agent. Both Hume and Mill considered, exclusively, instrumental rationality, while authors like Plato, (to some extent Hobbes), Smith, and Kant considered the existence of a stronger form of rationality – that is, one that also reasoned about which goals should be pursued. However, what about the formation of value – the valence of some goal – is that not also influenced the social environment? Smith argued, in hedonistic terms, that it was because utility was not solely based on individual utility but took into consideration social utility – that is, by anticipating how his actions might affect his peers. This perspective seems to imply that the intrinsic value of some reward is influenced, partly by how it may influence others belonging to one's social environment – those may either be physically present or imagined as generalized others or social norms – and partly by the well-being of others. In other words, it might be argued that the valence of some reward is influenced not just by the intrinsic value this reward represent to the individual, but also how this obtainment of a reward influences one's peers, and subsequently, how it influences the relation between the two agents. This aspect of utility theory, then, combines moral philosophical reflections on

social utility with individual utility theory. The concept utility is closely related with, intellectual historically, not just the hedonic foundation of rationality as well as morality, but also utilitarianism in Hume's approach, insofar as he seems to have introduced the concept and its relation with hedonism. Consider the two following quotations:

“It appears to be matter of fact, that the circumstance of UTILITY, in all subjects, is a source of praise and approbation: That it is constantly appealed to in all moral decisions concerning the merit and demerit of actions: That it is the SOLE source of that high regard paid to justice, fidelity, honour, allegiance, and chastity: That it is inseparable from all the other social virtues, humanity, generosity, charity, affability, lenity, mercy, and moderation: And, in a word, that it is a foundation of the chief part of morals, which has a reference to mankind and our fellow-creatures.” (Hume, 1748/1975: 259)

“The first circumstance which introduces evil, is that contrivance or economy of the animal creation, by which pains, as well as pleasures, are employed to excite all creatures to action, and make them vigilant in the great work of self-preservation. Now pleasure alone, in its various degrees, seems to human understanding sufficient for this purpose. All animals might be constantly in a state of enjoyment: but when urged by any of the necessities of nature, such as thirst, hunger, weariness; instead of pain, they might feel a diminution of pleasure, by which they might be prompted to seek that object which is necessary to their subsistence. Men pursue pleasure as eagerly as they avoid pain; at least they might have been so constituted. It seems, therefore, plainly possible to carry on the business of life without any pain. Why then is any animal ever rendered susceptible of such a sensation? If animals can be free from it an hour, they might enjoy a perpetual exemption from it; and it required as particular a contrivance of their organs to produce that feeling, as to endow them with sight, hearing, or any of the senses.” (Hume, 1779/1947: 206)

Plato identified a semi-strong version of rationality which had as its primary aim the balanced pursuit of one's desires – or overall wellbeing for the single individual as based on integrity and authenticity. Elevating the aim of overall wellbeing to the level of society enabled Bentham to push his conception of rationality past instrumental rationality (Gaus, 2008: 13). According to Bentham (1780/2007: 3), an action or a goal is irrational if it is against the total happiness of those affected by the choice. Bentham argues on the matter of hedonism:

“Nature has placed mankind under the governance of two sovereign masters, *pain* and *pleasure*. It is alone to point out what we ought to do, as well as to determine what we shall do. On the one hand the standard of right and wrong, on the other hand the chain of causes and effects, are fastened to their throne.” (Bentham, 1780/2007: 1, italics in original)

The rational agent should, then, take into consideration the material consequences of her actions in terms of hedonic experience – that is, she should in her process of deliberation take into account the expected pleasure (or happiness) of the parties involved (Bentham, 1780/2007: 71). What is important here is Bentham argument that only material consequences should be considered:

“Of the consequences of any act, the multitude and variety must needs be infinite: but such of them only as are worth regarding. Now among the consequences of an act...such only...can be said to be material as either consists of pain or pleasure, or have an influence in the production of pain or pleasure.” (Bentham, 1780/2007: 70)

Bentham’s conception of material in this quotation seems to explicate some important qualities about the causal relation: on the one hand, the result of some action produces within its target an emotional experience, however, not as a subtle mental emotion but as an externalized state. It seems here that Bentham is conceptualizing causality, in a positivistic sense, as the causal reaction between two objects – almost as the transference of energy. That is, the direct connection between two objects in terms of cause and effect, although he did recognize that some actions only indirectly generated this material as an antecedent of other actions (Bentham, 1780/2007: 70, f. 2).<sup>25</sup> In addition, this way of conceptualizing consequences also seems to have influenced the modern understanding of incentives in economics as some external force which causally generates within the agent an emotional experience corresponding to the action. Bentham even recognized these as the means of a legislator (Bentham, 1780/2007: 70). According to Bentham (1780/2007: 30-31) the consequence of an action should be determined by a hedonistic calculation (in modern economics often called a cost/benefit analysis) of pleasure or pain – the rational action then is one which maximizes the pleasure to the individual, the moral action is one that maximizes the total sum happiness, and since most actions has externalities the rational individual should take into account not only her own happiness but the total sum happiness – that is, the pleasurable or painful consequences of each affected individual:

*Individualistic hedonic calculation:*

---

<sup>25</sup> This rather naturalistic account of causality is also present in the work “Mathematical Psychics” by F. Y. Edgeworth, and in this work he describes utility as physical energy.

Pleasure (Intensity x duration x probability x nearness/remoteness) > Pain (Intensity x duration x probability x nearness/remoteness) = the action has a good tendency upon the whole hedonic value

Pleasure (Intensity x duration x probability x nearness/remoteness) < Pain (Intensity x duration x probability x nearness/remoteness) = the action has a bad tendency upon the whole hedonic value

*Total sum hedonic calculation:*

$\sum$  Pleasure (Intensity x duration x probability x nearness/remoteness) >  $\sum$  Pain (Intensity x duration x probability x nearness/remoteness) = the action has a generally good tendency

$\sum$  Pleasure (Intensity x duration x probability x nearness/remoteness) <  $\sum$  Pain (Intensity x duration x probability x nearness/remoteness) = the action has a generally bad tendency

The hedonic calculation seems then both to have been inspired by Hobbes' description of the relation between deliberation and expected utility, but also by the moral philosophy of Smith because, it takes into account the anticipated pleasure or pain of those affected, albeit it appears to be rather like *cold* moral calculation than a *warm* sympathetic relation as described by Smith. Additionally, one may also question the focus on consequences because does this not imply that the only antecedent of action is the anticipated utility and not that agents may be or feel themselves obligated to do something – that is, is it non-rational (or even irrational) to refrain from doing something not because one fears the consequences of this, but entirely because it is against one's principles or the governing social norms. The latter has already been discussed in relation to Smith's moral philosophy and the answer was that it was not. Nevertheless, one can question the foundation of this hedonic calculation because it seems to argue that the rational agent only does something if he believes it is within his or his peers' interest. It seems, of course, obvious that social agents do not just act; they most often have a purpose with doing what they do but is this purpose always material in its manifestation, that is, do social agents not also perform symbolic actions - actions without (or little) practical effect, and if they do, is it then to be considered irrational?

Furthermore, one may also question the complexity of such elusive concepts as pleasure and pain because can pleasure not also sometimes cause pain? Bentham, of course, was aware of this complexity like his predecessor Plato, that sometimes pleasures can be painful sometimes because of conflicting desires or due to one's values or the very nature of hedonic object of experience. Bentham also argues, as already showed, that the legislator should take this into account.

Utility, Bentham (1780/2007: 2) defines, closely resembling Hume's conceptions above, as that quality of an object which produces pleasure or avoids pain. Utility as such is considered the primary motive of rational and social action. Motives, as such, as long as they adheres to the principle of utility has no other qualities than their hedonic experiences, that is, in juxtaposition to Plato goals cannot be right and wrong in relation to some pure form of good or bad because they can only take on them the values imputed from hedonic experience, and as such pleasure is good, and pain is bad (Bentham, 1780/2007: 102). Indeed, Bentham's utilitarian perspective can be identified as a semi-strong form of rational choice; however, in this original form at least, it seems that there is no difference between the pleasure following from one's basic desires and those following from the satisfaction of some higher order of desire. This particular problem was also remarked by Mill (1883/2001: 10, 11) who argued that in opposition to Plato and Kant, true desires should spring from the experience of pleasure or happiness and not from some moral obligation. That is, the difference between higher and lower pleasures should not be turned into a distinction between value-systems; rather, higher pleasures are those which qualitatively bring the agent more pleasure than lower ones. Furthermore, one might distinguish between simple and refined pleasures, although one might rather quickly experience that simple pleasures are just another term used to describe the amount of hardship which goes into bringing them about – in a similar vein, mental faculties cannot be applied as a qualifier, insofar as the might solely enhance the level of complexity of experience without, actually, having any *real* influence (Mill, 1883/2001: 13).

Additionally, it also seems that Bentham solely believed that individuals had one master goal which any other sub goal needs to be related to. It seems, then, that there is an important difference between having happiness as master goal and having integrity

as one's master goal. While the former seems to argue that everything can be reduced to a pursuit of happiness, the latter argues that other goals might very well exist and what matters is that one stays true to oneself – that is, she might give money to a beggar because she holds a principle permitting her to do that or because empathy or indignation, and not because it makes her happy or to ease her bad conscience. Moreover, one should also question the principle of utility as the pursuit of total sum happiness because; on the one hand, this demands that such a thing exists and that it can be measured; and on the other hand, if it is to be measured it needs to be objective (or made objective) – that is, it needs to be separated from the agent without losing its meaning, thus it needs to be expressed on a cardinal scale. Even though experimental psychology has made some interesting result it is still at an early stage in its development of an objective concept of happiness and hedonic experience in general – I return to this in chapter V.

Bentham proposed, as already explained, a cardinal utility theory. In modern game theory, however, cardinal utility is often associated with the approach developed by John von Neumann and Oskar Morgenstern in their work “Theory of Games and Economic Behavior” (1944). Von Neumann and Morgenstern (1944/2004: 17) take as point of departure ordinal structured preferences and transform them into a cardinal utility function by using probabilities. In that sense, they propose a particular understanding of actions as lotteries due to the uncertainty of their outcomes. Assume that Ego has complete preferences between A, B, and C, and strictly prefers:  $A > B$ ,  $B > C$ , and by the axiom of transitivity it follows that  $A > C$  (von Neumann and Morgenstern, 1944/2004: 18; Elster, 2007: 197). Assuming strict preferences also implies asymmetry, that is, Ego strictly prefers A to B, however not,  $B > A$  (von Neumann and Morgenstern, 1944/2004: 18; 27). These four axioms are properties of ordinal utility theory as well. Neumann and Morgenstern (1944/2004: 18) then argues that given these preferences and the imputation of probabilities (making it a game with risk) enables a transformation of the utility function. They add four axioms to the ordinal axioms. The axiom of *continuity* argues that if Ego can choose between B with the probability 100%, and A and B both with the probability 50%, a game must exist which makes Ego indifferent between choosing A or playing the game with risk (Elster, 2007. 197). The sixth axiom implies that if Ego is to choose between two games one with A ( $p = 0,7$ ) and C ( $p$

= 0,3) and another with B ( $p = 0,7$ ) and C ( $p = 0,3$ ), he should because of preferences over (A, B, C) prefer the first game to the second game. The seventh axiom states that given two games one with A ( $p = 0,7$ ) and C ( $p = 0,3$ ) and another with B ( $p = 0,3$ ) and C ( $p = 0,7$ ), Ego should prefer the first game to the second game. And the eighth axiom states “...that it is irrelevant whether a combination of two constituents is obtained in two successive steps...or in one operation” (von Neumann and Morgenstern, 1944/2004: 27) – that is, two successive games can always be reduced to one (Gaus, 2008: 42).

Even though utility theory, as such, is not a model of rational choice, and has nothing to do, whatsoever, with utility maximization, as it is sometimes argued, because, as such, it is solely a model of *relative* preferences over objects, goals, situations, actions etc., based on some exogenous or, better perhaps, underlying structure of preference (Gaus, 2008: 32, 53). Utility theory is a formalization of preference theory, however, it is contrary to the logic of the model to apply it as means of assessment on rationality and irrationality because, as already argued above, an individual can choose not to pursue her first-order preference ( $A > B$ ) because she hold some second-order preference ( $B > A$ ) – she may desire to smoke a cigarette but refrains from doing it because she also desires to stop smoking. This agent is still, at least in the perspective developed in this text, to be considered as rational. Moreover, the individual who in one distinct context prefers A over B and in another prefers B over A, is not to be considered as irrational in regard to the axiom of transitivity because it follows the axiom of completeness. Above the Stoic virtue of constancy has been identified in both the philosophy of Montaigne and Hobbes, this virtue resembles to some extent the axioms of completeness and transitivity, although as such this virtue was aimed at the pursuit of true goals and one’s true nature, and not as an assessment of whether it is inconstant (irrational) to prefer oatmeal over cornflakes on Monday, and cornflakes over oatmeal on the following Tuesday. The axioms of completeness and transitivity have a functional value as an ideal conception of preferences, and as such it is an abstract simplification of the complexity of reality aiming, solely, at making it possible to model behavior. However, as a normative concept of rational choice they are problematic because it would imply that across contexts (spatiotemporal frames) the individual would have some idea of her preferences over (A, B, C), and that she always framed these choices

in a successive and like manner;  $(A > B)$ ,  $(B > C)$ , and  $(A > C)$ . This way of reasoning is close to the assumption in economic rationality on stable preferences (Gaus, 2008: 38). As identified by theorists on bounded rationality, choices are often made sequentially – that is, at one time they are facing a choice between A and B, and at another time facing a choice between B and C - just as humans are bounded by their limited experience – that is, they may never have found themselves in a situation facing a choice between A and C. Thus, choosing A over B in one context and B over A in another is only irrational if stable preferences are assumed (March, 2006: 204; March and Olsen, 1975: 150; March, Sproull, and Tamuz, 1991: 5).

Notwithstanding the above, it is easy to see why this conception of human preferences is so appealing because it induces logic. Gregory Bateson’s theory on the logic of learning (and the learning of logic) and its application to studies on schizophrenia seems to be built on a similar understanding of preferences because to people suffering from schizophrenia the same object may both be desirable and undesirable, an action might be both right and wrong – this double bind is contrary to the logic of preferences in learning in which one and only one action present itself as right. In that sense, the axiom of transitivity is useful in modeling rationality, as far as the actions of the rational agent are considered to be intelligible and rationality as decisive.

The axioms following from cardinal utility theory can, of course, also be questioned because, as Gaus (2008: 44) argues on the axiom of continuity, if one is to choose between A, B, and death it seems questionable that a situation exists in which the agent is indifferent between having A and getting B or having A and being dead. The two following axioms may be questioned. Consider the two games, (A,B) and (C,D), between which the agent is to choose:

<b>Options</b>	<b>p = 1</b>	<b>p = 89</b>	<b>p = 10</b>
A	100	100	100
B	0	100	500
C	100	0	100
D	0	0	500

In the first game (A,B) the agent might either choose to get 100 with certainty or gamble. According to the sixth and seventh axiom the agent should base her choice between A and B on her preferences for reward and the probabilities attached to these. Why should the agent choose option A over B? One explanation might be that because A and B have equal rewards if state 2 is realized the agent focus on the 1% probability of obtaining 0 in B compared to the 100 in A, and the 10% probability of gaining 400 more by playing B instead of A. In that sense, she might consider, in accordance with the axioms, the probability of gaining 100 in option A as 11% compared to the probability of gaining 500, 10% (Gaus, 2008: 46). Agents choosing option A should also choose option C over D because the probability of gaining 100 is 11% compared to the *only* 10% probability of gaining 500. However, agents have a tendency to choose D over C (Gaus, 2008: 47). Does this, then, imply that the agent is irrational? No, what it shows is that to understand the complexity of the choice one has to take into consideration false beliefs, limited computational skills and understanding of probabilities, and the underlying preference structure (Elster, 2007: 209).

### 3.3.7 Jon Elster and Donald Davidson

Following Jon Elster's (2007: 209) conception of rationality, one might argue that rationality is subjective in nature, because choices cannot be understood, as argued above, solely with reference to preferences over consequences. Rather, one should focus on the internal causal processes which produced the choice. This particular understanding of rationality, then, implies that one should not take beliefs and preferences as exogenously determined, rather one should perceive these as endogenously produced within the agent, albeit not intentionally, just as one should allow the agent to be placed within a particular situation which calls forth within her certain motives on which she is to act (Elster, 2007: 183; Merton, 1957: 110). On the one hand, then, Elster's (1977; 1989b: 17) critique of rationality is a critique of the blind faith in reason displayed by philosophers such as Kant; and, on the other hand, it is a critique of the conception of beliefs and preferences in economic rationality as exogenous constructions, insofar as this conception does not capture the influence of the context – internal as well as external (Elster, 1983; 1999). In regard to the former critique, it might be argued that rationality is self-defeating in some cases due to the limits of the will – the lack of external and inter-

nal control, see the discussion in the section on Kant. Furthermore, just as it is irrational to believe that one gain complete control of one's desires, urges (intrapsychic), produce actions which solely have good consequences, or interactional as in cases when the agent believes that her actions might cause other agents to act in a similar way (for example the Kant's categorical imperative), it is also irrational to believe that one through rationality can reach optimal decisions which maximizes the happiness of all parties concerned, insofar as the very act of being rational may undermine its result, as in the case of child custody disputes when the prolonged process following from the search of the optimal decisions ends up doing more harm than good and all parties were better off just tossing a coin and deciding at random what to do – that is, in some cases one's conscious attempt to bring about some particular end may be self-defeating (Elster, 2007: 120; 1989a: 192; 1989b: chapter 3).

In regard to the latter critique, Elster (2007: 165) argues that instead of preferences one should consider opportunities – the opportunities perceived by the agent – in the sense, that the former might be associated with identifying the most desired object, whereas the latter is associated with identifying the possible choices open to the agent, a move which enables Elster to explain the connection between macro and micro level, insofar as the opportunity set is determined by physical, psychological, legal, social, economic, etc. constraints, while the actual choice, within this opportunity set, is influenced by other factors such as the situation, social norms, etc. Opportunities are, then, related with the structuring of the modal verbs, “I can”, “I am allowed to”, “I will”, etc. For example, in the case of adaptive preferences – a matter I return to below – opportunities shape preferences (opportunities  $\rightarrow$  preferences), whereas in other cases, preferences might shape opportunities (opportunities  $\leftarrow$  preferences), as in cases related with character planning or weakness of will (Elster, 1983; 2007: 176). From this it also seems to follow that opportunities are solely produced externally nor internally, but emerges in the interaction between these – that is, the opportunity set faced by the is influenced by her previous experience (what she has been able to) in terms of, for example, physical accomplishment, just as it is influenced by her virtual memory of experience and the beliefs or narratives which these have been transformed into - narratives which, subsequently, comes into govern how she perceives herself, her abilities, her future opportunities in terms of the aspirations she hold, etc. (Elster, 1983: 123). The

agent's opportunity set, in other words, covers both her desires in terms of goals and preference structuring, and her preferences over different courses of actions – instrumental preferences. As a consequence of this, an agent holds preferences over what she desires, just as she holds preferences over how these desires may be brought about – actualized – instrumental preferences which cannot be reduced to means-end reasoning. As I will argue below, desires as well as instrumental preferences are centered upon information or evidence collected by the agent – her experience – and as argued above, in the section on Plato, not all information can be evaluated on the binary distinction between true and false experience. Opportunities produce the context – the internal and external context – of the decision, but the factors which goes into the production of the opportunity are not fixed across situations which implies that the context faced by the agent changes as the situations changes – in that sense, desires and instrumental preferences shifts as the context of the decision shifts (Elster, 2007: 184). This particular description of the interaction between the individual and the situation produces a particular understanding of personality as non-reducible to, more or stable, stable tendencies or dispositions, because situational characteristics may be more predictive of action than the attribution of disposition (Elster, 2007: 186). In that sense, behavior may be better explained based on the relation between the particular situation and the internal/external context – the agent-specific relation between the situation and the agent's behavioral propensities, thus personality can be reduced to a set of different response tendencies in particular situations (Elster, 2007: 188-189). The conception of the self which emerges is, then, fragmented. Although, Elster might be correct in his critique of essentialism, it can also be argued that in addition to this, the agent may also hold private norms on how to act, just as she might hold moral norms on how to interact with other agents, action cannot, as such, be reduced to *mere* behavior – at least not without reducing the possibility agency. These norms need not be rational on the contrary they can be irrational and sometimes inappropriate, but reflects the agent's prior decision on what is right and wrong. This can also be related with a third critique of rationality put forth by Elster (1989b: 36) regarding rationality as a principle of action and as a motive. Action and interaction cannot necessarily be reduced to neither rational principles regarding rationality of moral principles nor instrumental reasons, insofar as some principles of action, social norms, may not be endorsed because of their intrinsic rationality or their

ability to bring about desirable consequences, but solely because they are perceived as appropriate in a non-normative sense or as applicable in the situation and, as such, they come to guide action as non-consequentialistic principles of action not directed at how one best reaches one's goal, but solely as perceived socially legitimate principles on how to act, see chapter VIII for a further discussion of social norms. In a sense then, the agent may have reasons to endorse a particular social norm, but these reasons are not, as such, related with the principle itself, hence social norms are not reducible to reasons intrinsic to the principle itself; rather, a social agent may apply some principle because she fears how other social agents might react if she does not adapt her action to this norm (Elster, 1989b: 100). In that sense, the agent's reason to endorse a particular course of action may not be based on reasons directly related with the content of the principle (intrinsic causally related); rather, external reasons related with the transgression of the principle must be identified.

To understand Elster's (2007) conception of rationality as being a product of an internal causal process, one has to take into account Donald Davidson's (2001) idea about *rationalization* as a causal explanation which seems to have inspired Elster (1983: chapter 1). According to Davidson (1963/2001: 3) actions are based on reasons and these reasons rationalizes the action if and only if this reason enables one to understand *why* the agent chose this particular action, that is, her desires and beliefs which, then, becomes primary reasons of action (Davidson, 1963/2001: 4). If these are to be a primary reason, the *logical relation* has to hold that it states *why* that particular action was desired at a given event (a particular explanatory context), and the belief stating that this particular action given the event is desired (Davidson, 1963/2001: 5; 1978/2010: 173). In that sense, a primary reason contains a *why* expressed in its intentional form "*I wanted to...*", and a belief expressed in its explanatory form "*because I...*" (Davidson, 1963/2001: 6, 7). Here, Davidson (1963/2001) also infers the assumption that desiring an end also makes the agent desire the action necessary to bring about this end – in that sense, the means are not desired because of their instrumental quality, as means to an end, but are so because the end is desired. Consider the following sentence:

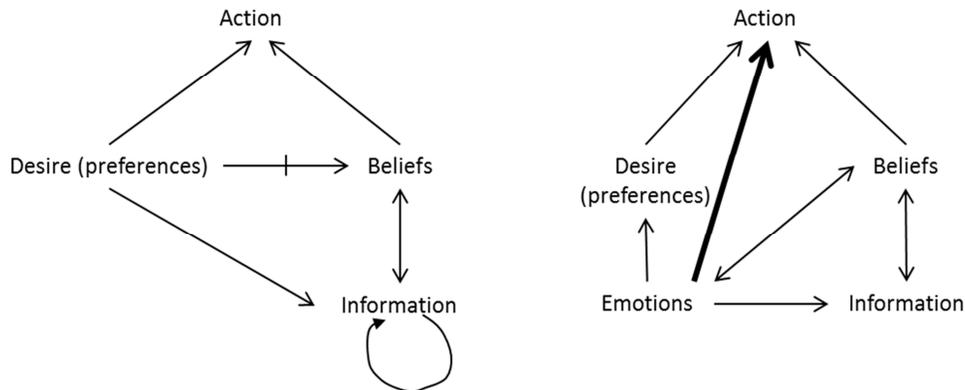
“The man stopped his car at a pedestrian crossing.”

This sentence describes an action, but it does not describe why the man stopped his car – a full account could be:

“The man stopped his car at a pedestrian crossing because he did not want to hit the woman crossing.”

This sentence explains the action in a manner which makes the primary reason “...because he did not want to...” the cause of the action – the *causal relation*. The explanatory context, here, contains both the action and its reason, just as it demonstrates a causal linkage which enables explanatory equivalence between cause and explanation (Davidson, 2001: 10-11). Indeed, as this example also demonstrates, one cannot explain an action by referring, solely, to its place within a larger causal pattern of external events which explain as little as “You stepped on mine toes” justifies “So I step on yours” (Davidson, 2001: 8, 12, 15). Elster’s (2007) theory on rationality can, then, be understood as a causal theory on the production of perceived rational actions, insofar as he assumes that social agents are subjectively rational, thus what needs to be explained is the formation of preferences and beliefs. Consider the two causal diagrams below; the one on the left hand side represents the causal process implied in *traditional* models on rationality, whereas the one on the right hand side represents the one explained by Elster above:

Figure 8: The production of action (Elster, 2009b: 15, 41).



The causal diagram illustrating the more classic approach to rationality, the one Elster (2007: 191; 2009c: 16, 23, 27) argues against is centered on three kinds of optimality:

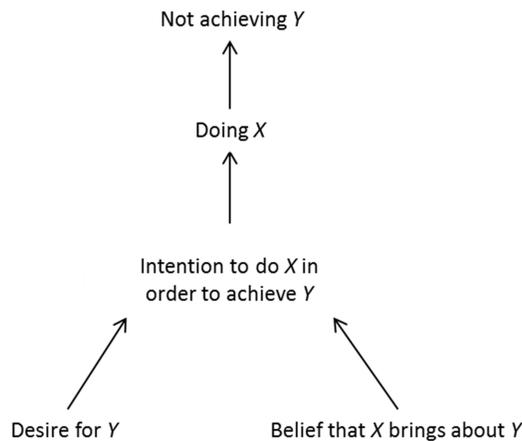
(1) instrumental rationality, that is, given the belief held by the agent about options and consequences of different actions available to her, she is to choose that which best optimizes her desire – represented by the arrows from beliefs and desire to action; (2) the beliefs held by the agent must be optimal in the sense that they are unbiased – based on proper (relevant) and an optimal amount of information – and well-founded, understood in the sense that they are caused in the right way; and (3) the amount of information collected – the resources invested – depends on the level of desire (importance), just as it depends on the agent’s judgment on when she has sufficient grounds for her belief – the self-referential arrow below illustrates that gathering more information sometimes generates a higher need for more information or alters the meaning of what is already collected. The arrow running from desire to belief is crossed-over because it represents motivated belief formation which is irrational.

The causal diagram on the right hand side illustrates the subjective nature of rationality, just as it also illustrates why irrationality occurs. The thick arrow running from emotions to actions illustrates that, occasionally, social agents do act directly on their emotions, as it happens in cases of weakness of will. Desires are not, as beliefs centered, on rigid criteria on rational formation, inasmuch as the primary criterion for rational desires in coherency stating that one cannot desire A and B equally if these are mutually exclusive and other occasions of irrational preferences are self-defeating desires (when the desire itself causes the event not to occur), hyperbolic preferences, and the belief that one by pure will can overcome something (Elster, 2009c: 18, 19, 23). Preference reversals may occur, as illustrated by the arrow between emotions and desires, as in cases where the agent at time-0 prefers A to B, at time-1  $A < B$ , and at time 2  $A > B$ , that is, the agent’s fear of rejection may make him refrain from speaking with a beautiful girl at time-1, and at time-2 he wishes that he had done it (Elster, 2009c: 42). At other occasions, emotions might also influence preferences by heightening the agent’s subjective probability of an event occurring, just as it might make the agent prefer early rewards to later ones (impatience); or to act now instead of waiting – that is, it increases the urgency with which the potential action presents itself to the mind (Elster, 2009a: 58, 63). This sense of urgency may also, indirectly, cause biased beliefs because it makes the agent refrain from collecting evidence (information) about the long term consequences of her actions, and as such it blinds or, at least, prevent de-biasing from occurring (the

arrows running emotions to information to beliefs) (Elster, 2009c: 43, 44; 2009a 68; 2007: 155). Finally, the arrow running from emotions to beliefs illustrates how the mind sometimes plays tricks on itself as when one really wishes something to be true and as a result come to believe it (Elster, 2009a: 67).

With this in mind, one may now ask when an action is irrational. Elster’s early authorship is devoted to answering this question and can be summed up in one sentence – irrational actions are *mistakes*. Irrational actions are mistakes, not however simple mistakes such as confusing one action with another or failing to bring about some desired result; rather irrational actions are actions internally *caused* in a wrong way (Elster, 1989: 19). That is, when the mistake refers to, for example, false beliefs being the foundation of an action, then the action is caused in a wrong way, because these wrong beliefs are intrinsic to the very production of the action. One cannot, for example, make oneself believe or fall in love at will insofar as these are spontaneous occurring states (by-products) and just trying to bring them about might undermine the success of one’s actions, these actions are in other words self-defeating as argued above (Elster, 1983: 45; 1989: 19). This might be illustrated as follows:

Figure 9: Self-defeating actions (Elster, 1989: 19).

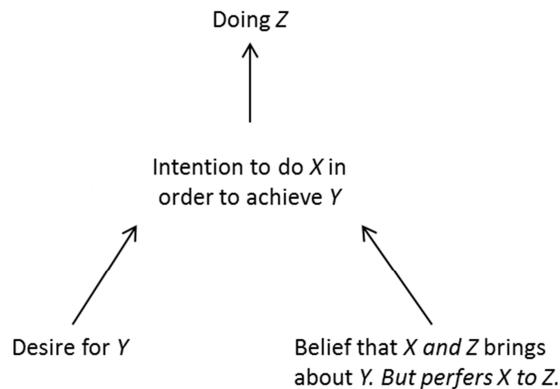


As this figure illustrates, the action in itself might not be irrational – *X* might be rational in itself, however, if it is taken to bring about a state which cannot be willed, then its prerequisites in terms of beliefs are wrong, and as such the action is irrationally produced. To comprehend the complexity of this, Davidson (1978/2010) once again can be of some assistance. Davidson (1978/2010) initiates his analysis of irrationality with the

statement: “Irrationality is a mental process or state – a rational process – gone wrong.” (Davidson, 1978/2010: 169).

According to Davidson (1978/2010), then, irrationality implies rationality, hence a paradox emerges. This paradox is due to the way actions are explained (Davidson, 1978/2010: 174). That is, the implied intentionality which emerges as a consequence of the two relations explained above, the logical and the causal relation. Irrationality must, in other words, refer, as Elster (1989: 2007) also argued, to a mistake or failure within the action’s causal pattern of production – beliefs, desires, intentions, emotions, ends, etc. – in terms of consistency or coherence (Davidson, 1978/2010: 170). If all actions are caused by reasons – even when arguing that the action had no reason, one is just arguing that it had none other reason than the one displayed in the action itself – then irrationality implies that the agent is, by being irrational, acting against her better judgment, and this problem has troubled philosophy on rationality since Plato. This can be illustrated as follows:

**Figure 10: Acting against one's better judgment.**



Plato and Kant solves this problem by inferring that such an inconsistency cannot occur because of internal consistency brought about by the existence of what Plato called the soul and Kant later called the identity, see the sections above. Hume, on the other hand, by arguing against this internal unity, stands at the other end of the extreme arguing that such an inconsistency cannot occur, because it would imply that someone else, besides the acting agent, was doing the choosing insofar as the action, here, cannot be contrary to reason. The problem, however, may not be the complexity which emerges from this state of contradictory reasons, insofar as this might be fairly easily overcome by weigh-

ing pros and cons against each possibility, the problem is according to Davidson (1978/2010):

“...although the agent has a reason for doing what he did, he had better reasons, by his own reasoning, for acting otherwise. What needs explaining is not why the agent acted as he did, but why he *didn't* act otherwise, given his judgment that all things considered it would be better.” (Davidson, 1978/2010: 176)

The problem, then, is not that from the perspective of another agent the acting agent acted in a manner which seemed irrational; rather the problem is that the acting agent is acting against her own principles. Consider the following story:

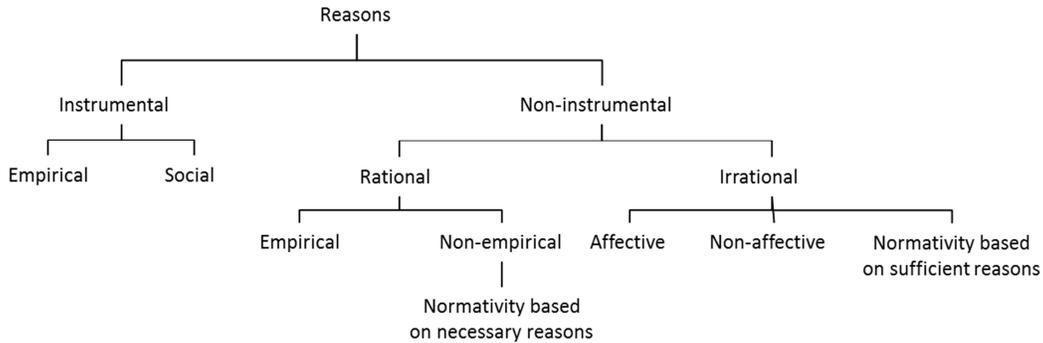
A man walking on the side walk stumbles upon a broken flask and almost cuts himself. To secure that no one else cuts themselves he kicks the flask onto the road. Walking home he comes to think about what if a car hits the broken flask. He decides not to go back and put the flask into a trashcan because it is too far. Nevertheless, he goes back and puts the flask into the trashcan.

By walking back and putting the flask into the trashcan, the man is actually acting against his better judgment, his decision that it was too far. In the first part of the story, he decides to kick the bottle onto the road instead of putting it into the trashcan implying that he, for some reason, preferred kicking it onto the road to putting it into the trashcan. In the second part of the story, the man decides against putting the flask into the trashcan because it is too far to walk back. In the third part of the story, however, the man's desire to put the flask into the trashcan overrides his previous decisions. The man's desire for putting the flask into the trashcan cannot be viewed as a reason for overriding his previous decision – a reason against his decision – and as such it is irrelevant as a reason in itself, and as such it violates the logical relation explained above. Had the man not decided against walking back, in the first place, his action would not be irrational insofar as it would not be against his own decision. Irrationality, in other words, implies, simply stated, a *hole* in the causal chain which produced the action, and as such irrationality does not imply incomprehensibility from the perspective of a second person and neither does it imply a lack of reason, because the man had a reason, he wanted to put the bottle into the trashcan, the reason was, nevertheless, not equivalent with the action he chose (a violation of the causal relation).

### **3.3.8 Raymond Boudon and the Positive and Normative Dimension of Rationality**

The rational action may, then, be incomprehensible and obscure, but that does not imply lack of reason (mindlessness) or irrationality, thus every action can be referred back to an acting agent, and must, as such, be understood on her terms – with reference to her reasons (Weber, 1922/1968: 5). Following this *Weberian* understanding of rational action, Boudon (1996: 124; 2006: 164; 2011: 34) develops his, sociologically founded, theory on rational agency stating that rationality means nothing but assuming that the agent is acting, believing, etc. based on set of reasons which she holds to be individually strong and combinable with one another, just as this set of beliefs must be defensible to a possible alternative set, that is, no other set must be preferred to the actual set. In that sense, an action is rational if: (1) it is based on reasons; (2) a rational action needs no further explanation of why it is endorsed – rationality is a self-sufficient reason; and (3) if it instrumental it is only so in a trivial sense – as enabling the bringing about some state. According to Boudon (1996: 124; 2001: 58), instrumental rationality creates too many black boxes when it explains actions as being chosen primarily as a consequence of their potential of bringing some desired state about, because if the agent hold a goal and believes that a particular course of actions will bring this goal about, instrumental rationality only explains why the agent chose the action most likely of bringing the goal about, and not why the agent, in the first place, believed that this particular course of actions would bring the desired state about. That is, instrumental rationality only provide a partial explanation by focusing on instrumental reasons, whereas the cognitive link associated with the believed causal connection between some action and some goal, the non-instrumental dimension is left out, because beliefs, as explained above, are viewed as exogenous. Moreover, not all actions can be explained with reference to instrumental reasons, insofar as some actions might also be purely non-instrumentally in nature – actions, for example, based on principles (Boudon, 2001: 61). The question is now, why does the agent endorse such a principle – like the cognitive *bridge* between the instrumental and non-instrumental dimension becomes, why the agent endorses such a causal theory. Consider the following figure:

Figure 11: Hierarchical taxonomy of some basic distinctions of reasons applied by Boudon and Weber.



As the figure illustrates, it is analytically possible to distinguish between instrumental and non-instrumentals reasons, the former being associated with what Weber (1922/1968: 26) calls goal rationality; and the latter being associated with what he calls value rationality. The instrumental dimension can be subdivided into an empirical and a social dimension. The agent’s choice of means with regard to the instrumental dimension may be centered upon empirical or social expectations, that is, the agent’s choice of means is conditional upon her empirical expectations about how non-human objects function; and upon her social expectations about how social agents acts and reacts (Weber, 1922/1968: 24). The non-instrumental dimension can be subdivided into a rational and an irrational category, the rational being that which is within bounds of reason and meaningful; and the irrational being that which is opposite – the incomprehensible and ungrounded in reasons (Boudon, 2001: 123). The rational dimension can, then, be subdivided into an empirical dimension and non-empirical dimension, the former being justified with reference to empirical evidence; the latter cannot being justified with regard to empirical evidence, but is, as Weber (1922/1968: 25) argues, based on the agent’s *conscious* endorsement of some value or principle. The irrational dimension can be subdivided into an affective dimension associated non-conscious reasons; a non-affective dimension associated with cognitive errors as the ones I will discuss below; and a “normativity based on sufficient reason” dimension which should be distinguished from the “normativity based on necessary reason” dimension. The former logical condition stating “if X, then Y” and the latter “Y if X”, the former stating that if X is true Y is also true, while the latter states Y cannot be true unless X is also true (Boudon, 2001: 174). A perhaps simpler method would be to distinguish between normative reasons

based on conscious reasons to endorse these, and normative reasons which are *just* held to be true in a universal sense (Weber, 1922/1968: 26). Weber (1922/1968: 25), of course, also argue that tradition may be a reason which in some cases can be rational, while in other cases, mindless repetition, takes on a more irrational nature. Based on this *Weberian* inspired system, Boudon (2001: 59) advocates the perspective that action should be analyzed as based upon rational reconstruction of motives which can be confronted with data. In a similar vein as Weber, Boudon (2001: 62, 98) argues that actions should be considered as meaningful to the agent, the meaning should be considered as the cause of the action, just as the meaning should be viewed as being based upon reasons – that is, the reasons the agent has to endorse it.

Boudon's (2001) point of departure is the non-instrumental rational normative reasons which he associates with cognitive rationality or *axiological rationality* based on *axiological beliefs*. An action is axiological rational if it is based on the endorsement strong theories rather than weak theories (Boudon, 2001:149). The endorsement of one theory over another is a non-instrumental act, thus the central question is not whether some action maximizes the utility; rather, the question is why the agent finds the theory to be defensible (Boudon, 2001: 65, 149; 2011: 36). An axiological belief is a belief which is based on strong reasons – reasons being a complex system of arguments which produces a theory, and as such the context of the belief (Boudon, 2001: 47; 2011: 39). That is, the agent holds X to be true if: (a) within the system of arguments X is arrived at; and (b) if X has no serious competition from other conclusions which could be arrived at based upon the same system of arguments. In other words, combined the two criteria states that the agent holds X to be true given her reasons – the belief is, then, grounded in evidence (in a logical, but not necessarily empirical sense), hence these beliefs obtain an inter-subjective claim on validity, insofar as it is believed that another agent given the same chain of arguments would reach a similar conclusion (Boudon, 2001: 68; 2011: 39). In that sense, the agent believe that due to solid reasons X is true, she also believes that another agent should endorse the same theory – thus, beliefs should be distinguished from preferences, inasmuch as the agent accepts that other agent prefers X to Y (Boudon, 1996: 128; 2001: 124; 2011: 40). The trans-subjective validity of the belief related with the agent's belief that another social agent would reach a similar agent can be associated with Mead's (1934/1992: 154, 155) notion of the *gen-*

*eralized other*, in the sense that the agent must not only take into account how she relates to others or how they relate to her, but also apply this internalized generalized otherness in the formation of her attitude towards social practice of a more general notion. Hence, this might also be viewed in accordance with Boudon's (2001) theory on the impartial spectator discussed above. The trans-subjective claim on validity should, however, not be confused with a claim on objectivity or truth, insofar as the theory by the agent is perceived to be true due to the intrinsic strength of the arguments, hence the agent may have strong reasons to believe it and believe that another agent given the same arguments would reach a similar conclusion without this meaning that the theory is objectively true (Boudon, 2001: 126, 128, 131). Values, like principles, are to resemble with conjectures on what is held to be right without being based on evidence, whereas factual beliefs, as argued above, can be evaluated drawing on the binary distinction false/true.

Boudon's (2001: 53, 68, 103, 111) claim is, then, that normative statements, like positive ones, are based on reasons. That is, an agent hold X to be good, right, legitimate etc. if she has strong reasons to believe this to be the case, and she believes that other social agents given the same arguments would reach a similar conclusion. The agent's reason for endorsing X is, then, not based on consequentialistic or ultimate principles, the former being associated with instrumental rationality, and the latter with irrationality.<sup>26</sup> Boudon, in other words, proposes a moral theory based on *Verstehen*:

“...when people believe that something is true, fair, legitimate, etc., the cause of their belief resides in the meaning of the belief to them, in other words, in the reasons they perceive the belief as strong.” (Boudon, 2001: 49)

What Boudon seems to be proposing is a moral theory based on a parallelism between normative and positive theories, in the sense that positive theories have a normative dimension, and normative beliefs have a positive dimension (Boudon, 2001: 53). Consider the table below:

---

<sup>26</sup> Like Elster (1989), Boudon (2001) argues against the Kantian conception of rationality, although the argument here is based on a critique of the character of the claims, whereas Elster argued against Kant's faith in rationality.

Table 1

Justificatory dimension	Proposition
<b>Normative</b>	“I hold X to be good, fair, right, etc.”
	“I believe that other people also hold X to be good, fair, right, etc.”
<b>Positive</b>	“I have reasons to believe that X is fair, good, right, etc.”
	“I believe that given the same arguments, a generalized other would reach the same conclusion as I, X is good, fair, right, etc.”

Boudon (2001: 110) views the second normative proposition which states that the agent also believe other social agents to hold the theory, as the point of departure for moral sentiments – however, to distinguish them from affective reasons, one might view them as normative expectations. This also entails that the emotional intensity accompanying a potential transgression of these normative expectations is correlated with the strength with which the agent believe these normative expectations to be endorsed by others, and the strength with which she believes others to reach to same conclusion (Boudon, 2001:140). According to Boudon (2001: 151), this parallelism enables him to escape the *naturalistic fallacy*, because the normative propositions cannot be derived from the positive ones – one cannot derive normative proposition “ X should be endorsed” from “X is good” without presupposing the normative proposition “X is held to be good”. That is, at least one of the beliefs must be axiological in nature – hence, one must presuppose the existence of some principle from which one can derive a set of consequences, which then can be evaluated, and based on this evaluation one can evaluate whether X is indeed good. The process, then, mimics Popper’s process evaluation in the sense that the process starts with a *bold conjecture* which then needs to be tested against evidence, logical as well as empirical. Moreover, the perspective proposed is also similar to the one proposed by Mill who argues that to understand the strong emotional reactions associated with moral transgressions, one must presuppose the existence of some principle which have been violated – as Elster (1999) would argue – emotions are invoked by beliefs.

In that sense, one might view Boudon's (2001) theory on social rationality as a restatement of Kantian social rationality, although unlike Kant, Boudon adds a positive dimension. That is, the claim passed by the social rational agent is no longer just a claim on unconditional acceptance, it is a claim backed by reasons, and as such it can be justified. The principle can be cognitively justified, insofar as it is based on certain reasons which are acceptable to the agent self. Boudon, then, seems to add to this discussion that the rational agent does not solely hold a principle; she has reasons to do so – a non-instrumental reason. In a similar manner, it might also be argued that one should not consider deviations from some standard as irrational; rather one should try to inquire into the foundation – implying that instead of attributing some irrational motive to the agent, one should enquire into the foundation of the judgment, just as one should presuppose that the agent as being in possession of a rational competence – that is, the judgments made by the agent are not just passed, they are based on certain reasons and can, at least at some occasions, be supported by arguments allowing one to assess the foundation. In that sense, Boudon's *Verstehen* based morality follows, to some extent at least, the lines of the Platonic-Kantian approach to moral philosophy with its focus on cognitive and evaluative rationality as its foundation. The validity of the judgment, however, is not as argued by Elster (1983: 16) simply a matter of internal causality, it is also influenced by public opinion and how the agent tries to pass judgments which are not purely egocentric in nature (produced in the continuum between “how I see myself” and “how others see me” related with esteem/self-esteem) it is also how the social agent tries to take into account, when passing her judgment, the opinions of these others on more general matters. That is to argue that these judgments are externally caused – solely that they must be believed to be externally defensible, but not necessarily in accordance with public opinion. From this it seems to follow that a particular belief is accepted if the reasons one have for holding it are defensible in the light of existing, possible, counter-reasons – this, then, seems to imply that a particular belief is held as long as no higher preferred counter-belief is produced, thus, in a vertical perspective, a particular distribution is accepted as long as no strong counter-reason is produced against it (Boudon, 1996: 144; 2011: 41). That is, one needs not, necessarily, hold the same values as the one's expressed in a particular compensation system in order to accept it, one only need not to have any reasons to find it illegitimate and, as such, it is rationally accepted

as long as it has no strong reasons against it. Acceptance, in other words, need not be based on external and irrational reasons, such as power-relations, dependency, socialization, etc. but can be rationally founded in cognitive rationality.

Above I argued that humans have the capacity or competence to pass rational judgments based on the internal logical production of defensible judgments. In addition to this, it might also be argued that humans have the capacity to transcend the immediate present and produce judgments of a more general nature, and that this capacity is influenced not just by the social agent's capacity to take the position of the other, but also by her capacity to take into account more generalized accounts of what is right and wrong. On the one hand, the judgment produced is a normative statement, which is backed by positive reasons. Whereas capacity relates to ability, competence relates to the active engagement with the real world from which these abilities emerge. Rationality as a competence, then, relates with social agents' ability to produce valid judgments – judgments which are reasonable – understandable and acceptable – to the agent self and others. If this *Verstehen*-based morality is taken seriously, then, the moral judgment is based on reasons held by the subject passing the judgment, but these reasons need not solely be subjectively justifiable, insofar as they also are passed with a claim on acceptance based on the acceptance of reasons expressed in the positive dimension – the dimension of internal causality. Hence, acceptance cannot be, as argued above, reduced to matters related with power nor matters related with collectively held values, insofar as neither of these explanations addresses the positive dimension of judgment (Boudon, 1996: 126. In that sense, a judgment in order to be inter-subjectively assessed needs to be formulated in general terms which can be evaluated by others. Understanding moral judgment, in other words, becomes a matter of transcending the immediate experience and enquiring into the internal/external context of the judgment, in an abstract sense, the construction of the situation and the transformation of this into a judgment of the particular based on general terms – the abstraction from the particular to the general level. To address the particular through the general level should, however, not be confused with universality, insofar as this would require that the validity of the judgment should also be assessed in a universal sense – that is, what is argued here is not that social agents have the capacity to pass universal statements which are universally valid; rather, what is argued is, simply, that social agents have the capacity to pass general judgments

based on abstraction from the concrete or the particular to the general. In that sense, a judgment formulated on general terms has its base in the concrete – the situation – but through abstraction to a general level, the social agent becomes able to transcend the particularity of the situation and address it on general terms. Rationality is, then, tied to morality through its conceptualization as a competence to self-critique – a capacity to assess the validity of one’s judgments and the reasons one has to endorse them in the light of other systems of reasons which might put forth arguments against one’s judgments. A moral judgment, passed in accordance with this, must be internally as well as externally defensible and, as such, one is committed to a meta-principle stating that one should only hold judgments based on non-defeated reasons. From this perspective, then, the judgment is endorsed simply because it is rational – there are no strong reasons against it, at least none which the agent holding the judgment can produce which is not to argue that no reasons exists. By engaging with the real world and other social agents, the social agent may change her perception because her reasons are defeated by other social agents’ arguments and, as such, the moral judgments passed by the agent also takes on a social dimension stemming from the fact that she is influenced by the perceptions held by other social agents – social critique, external in origin, may then also be taken into account.

### **3.4 Tentative Conclusions**

The aim of this chapter has been to discuss the foundation of fairness as centered upon a distinction between rationality, irrationality, and social rationality in order to identify the philosophical foundation of vertical as well as horizontal fairness. In the first section of this chapter, I distinguished between the instrumental, the relational, and the moral approach to organizational fairness. According to Folger (1998; 2012), the moral approach fills a theoretical gap in organizational fairness theories, insofar as it re-introduce morality into fairness theories. Fairness theories have, according to the moral approach, lost their moral roots and focus more on rationality – either in an instrumental approach as trying to gain control of the outcome or in a socioemotional sense as a matter of inclusion/exclusion. According to Folger (2012), the metatheoretical foundation of these theories is rationality, just as these theories, originally, emerged in opposition to

theories based on economic rationality. In that sense, Folger tries with his deonance theory to identify a non-rational approach to fairness theory based on a deontic approach to morality. One might here argue that deontology neither in a Kantian sense, Benthamite sense, nor in a modern approach tries to liberate deontology from its rational foundation, insofar as rationality is closely related with the process of reasoning on which moral judgment is centered. Hence, morality without some kind of rational basis is, at least, from an intellectual historic perspective a reduction, because rationality does not, necessarily, imply egoism. The assumption on rationality is, of course, not unproblematic, insofar as rationality may demand too much in terms of, for example, the social agents' cognitive abilities and the information available to them. Rationality, in other words, enters in, here, as the foundation of reasoning (internal causality), as related with motivation, and as related with the collection and processing of data. Besides confronting fairness theory with its moral and rational foundation, a second aim of this chapter has been to understand the moral foundation horizontal relations in organizations – that is, symmetrical employee-colleague relations as an extension of the vertical perspective.

Rationality cannot be reduced to instrumental rationality on means-end relations, insofar as the majority of the classical writers on rationality also advocated the existence of stronger forms of rationality – that is, rationality might also include reasoning about truth, goals, values, etc. This also implies that rationality have an instrumental and a non-instrumental dimension. The instrumental rational approach emphasizes that the goal is brought about in an effective manner; while the non-instrumental approach to rationality emphasizes that the goal being pursued is based on proper beliefs and expresses what is *truly* good to the individual, and that it is produced in an autonomous manner. The instrumental approach to rationality, in other words, reduces rationality to a matter of choices between different courses of action without enquiring into why the particular goal is pursued and why a particular action is endorsed as an effective means in bringing the goal about. In a Platonic (360BC/1961; 380BC/1961) sense, rationality becomes a principle self-governance, insofar as it becomes the very foundation of a life based on reason, just as it becomes the principle which enables, along with reason, the unity of the soul. From a Kantian (1781/1787/2007; 1785/2002; 1788/2004; 1793/1794/2009) perspective, rationality also becomes the principle of self-governance, just as it becomes a principle on social life through his idea about the kingdom of ends.

Like Plato, Kant focus on the social agent's ability to produce rational judgments and like Plato he also addresses the problems which emerge from impure experience and adds to this framework the influence stemming from the social world in terms of a focus on happiness and wealth – that is, the influence stemming from different value-systems. A rational action is one that unfolds the will, the action must be in accordance with the will, and the will must be in accordance with the principles of reason – in the sense that one must will the action and its consequences. Being a Kantian agent entails, on the one hand, responsibility insofar as the agent's will is the causal origin of her action – it implies an active engagement with the world – just as it, on the other hand, entails courage in the sense that the agent must be courageous enough to take upon herself the responsibility associated not just with self-governance, but also with producing actions which have an impact on the life of other social agents – hence, an action must never treat other agents as means to an end. Kant, in other words, expands the Platonic conception of morality from self-governance to include a reference to other social agents and, as such, he also expands on the Platonic notion of responsible. The will to engage with the world, with other social agents, to be moral, etc. is the motivation behind the action, thus the action has no object besides itself which it tries to bring about and neither is it externally caused. This notion of the will as the motivational point of departure is also found in Deci and Ryan's (2000) approach to human motivation. The autonomous will springs internally from the will itself and, as such, it is not caused externally, this, however, does not imply socially independence, insofar as it is solely in this state of autonomy that authentic life can unfold itself – authentic living and authentic social interaction, a matter I return to in chapter VIII.

Hume (1739/1985) wrote in opposition to the Platonic perspective that rationality was purely instrumental and related with bringing about some desired state – hence, if rationality is based on reason, reason is entirely the slave of passions, insofar as it have no goal besides these defined by desires. In a similar vein, morality is based on the emotional experience experienced when conducting some action; hence the foundation of morality is hedonic experience. Hume followed the in the steps of Hobbes (1651/1985) who argued that rationality was about the deliberate production of intentional states and, as such, rationality was entirely a matter of bringing about some desired goal. Hobbes however did not view reason as the slave of passion, insofar as he acknowledged that

excessive pursue of one's desires would undermine the primary aim of rationality, self-preservation. In a similar manner, if the agent did not take into consideration the consequences her actions might have on other social agents' life this might lead to a state of war and, as such, this would also be against the primary principle of rationality. In that sense, desires was constrained by two limits – one centered upon passions within reason and another centered upon passions within social acceptance or social equilibrium. Hobbes, however, did not believe that social agents could govern themselves without the introduction of some external contract to govern the interaction. Social agents' engagement in such a social contract was based on rational principles; hence a social agent would solely adhere to the contract as long as it was within her interest to do so. In a similar manner, Coleman (1990) argues that social life should be understood as the exchange of rights and obligations. To engage in a social contract implies that the social agents transfer some of their rights to self-determination to the collective in exchange of some claim on each other. This transfer of rights, albeit purely symbolical in nature and often purely expressed in the fact that social agents take into consideration other interests than their private interest, can also occur to non-human agents as it is the case with social norms.

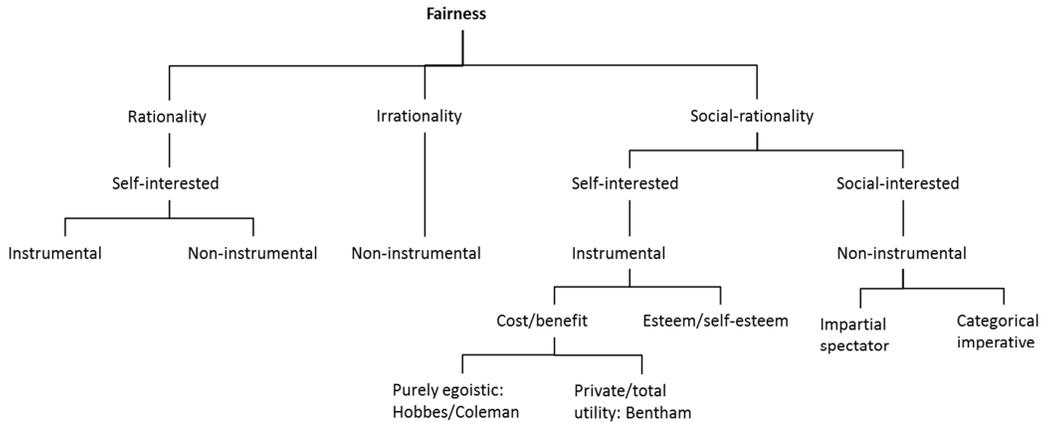
Social norms understood as something which governs social interaction has its intellectual roots in Hume's idea about social conventions. Unlike the Hobbes-Coleman approach which reduced it to a matter of costs and benefits, Hume argued that social agents adhered to these norms because of their social nature and the fact that their esteem/self-esteem was related with how other social agents perceived them. In opposition to this, Smith (1759/2009) argued that morality could not be reduced to cost/benefits or matters purely related with esteem/self-esteem, insofar as sympathy also implied that social agents had the capacity to take into account how their actions might influence other social agents by taking their perspective. Based on this, Boudon (2001) argues that it is this ability to take on an impartial spectator perspective on oneself which enables social agents to transcend the immediate situation and take into consideration other social agents' interests. Moreover, in combination with this ability, Boudon (1996; 2001; 2011) also proposes a model on rationality which takes the non-instrumental dimension seriously, because as he argues, rationality means nothing but acting on strong reasons. In a similar vein, a moral judgment is also based on strong reasons which the agent has

to endorse it and, as such, a judgment is rational if the agent has no reasons not to endorse it – it is not defeated by other value-systems and it is defensible. A judgment may be separated into a normative and a positive dimension, insofar as it is an expression of what the social agent holds to be good, right, fair, etc.; and a positive dimension related with the reasons the agent has to endorse the perspective. Accompanying this judgment is also the claim that other social agents given the same arguments would reach a similar conclusion. The Kantian tacit claim on respect is, then, transformed into a claim on trans-subjective validity. The claim on validity is, in that sense, no longer a claim on universal validity, but solely general validity in the sense that the judgment is justifiable and formulated in abstract terms – the ability to transcend the immediate situation is, then, a matter of the agent's ability to produce a judgment as an abstraction on the particular situation. This ability is enabled, as argued above, by the agent's ability to take on the perspective of the other and view herself through this sense of immanent otherness, a perspective which also enables the agent in producing judgments which are not entirely egoistic, insofar as it allows her take into account the other on his terms.

In relation to the three perspectives on fairness discussed above, the instrumental approach may be related with the Hobbes-Coleman approach, the relational approach may be related with the Humean approach, and Folger's perspective can be associated with the Plato-Kantian approach. In that sense, from a metatheoretical perspective, the three approaches can be associated with three different currents in rational philosophy. These three different perspectives can be viewed in opposition to the Benthamite (1780/2007) approach to rationality, although he followed Hume's approach, at least to some extent, insofar as he centered his approach on a utilitarian principle. Bentham, of course, did not solely focus on egoistic behavior, maximization of private happiness, because the moral action should maximize the total sum happiness. In contrast to the Hobbesian perspective focusing on regaining control, Bentham focuses primarily on choice and its consequences. The Bentham-Hume-Hobbes perspective leads to the conception of economic rationality, in the sense that inspired by Hobbesian instrumental rationality and means-end reasoning was introduced, just as it has been inspired by his assumptions on human anthropology. Inspired by Hume, economic rationality is based on the assumption that reason is the slave of the passions; and from Bentham it has taken the assumptions on consequentialism and the relation between preferences and

choice. Based on what has been argued so far, the following hierarchical taxonomy of fairness can now be illustrated:

Figure 12: Hierarchical taxonomy of the relation between rationality and fairness.



In contrast to the rational and socially rational action, an irrational action is not just an incomprehensible action or an action which deviates from some norm on what is considered as rational, just as it is not an action based on motives belonging to another social domain. An irrational action is an action which is caused in the wrong way – that is, it is centered on false or wrong beliefs; or goals which are not truly good. The irrational action, in other words, is contrary to one’s best judgment and, as such, it is not contrary to some *fictive* standard claiming universality, at least not in an objective sense, but solely as a matter of internal causality – entailing that the individual given her beliefs might claim generality, in the sense that she believes another rational agent given her beliefs would reach the same conclusion. An irrational action is one which is based on wrong or an erroneously foundation of reasons or it might also be, as just argued, in situations when the agent acts against her better judgment. Hence, an action which is contrary to reason, in the sense just described, is to be considered as irrational. An irrational action cannot be considered as objectively irrational, only subjectively irrational in the sense that it would not be contrary to the agent’s reason to act in a particular manner if she, given her beliefs, found this to be correct. Hume might then, at least to some extent, be right, although the action can be contrary to subjective reason – the agent’s best judgment. In accordance with the Davidson-Elster perspective, an irrational

action is one which is opposite to the agent's best judgment, insofar as this implies that the action cannot, logically at least, be derived from the system of reasons on which the judgment is based and, as such, the action is caused in a wrong way. Furthermore, Hume argued that an action based on desire cannot be irrational in itself, insofar as passions cannot be considered as irrational in themselves without assuming a false contrast between passions and reasons. Elster (1989a) argues in a similar manner that blind faith in the omnipotence of rationality is irrational, insofar as rationality, in a universal sense, is an unattainable ideal, because it presupposes too much of the agent in terms of, for example, internal causality (the limits of the will) and cognitive faculties. Firstly, then, an action is based on a cognitive irrational foundation if it is claimed to be universally rational; secondly, an action is irrational if it presupposes irrational beliefs about internal or external causality; and, thirdly, an action is irrational if it is caused in the wrong way – based on wrong beliefs and wrong values. From this it seems to follow that it is irrational to believe and will something which is not within one's power – that is, it is irrational to will that something happens which is beyond one's power, just as it is irrational to believe that by doing something one can make another thing occur, while the former can be associated with internal and external causality, the latter can be associated with magical thinking. In a simplified manner, one might also argue that while the choice between the shortest road between point A and B and the beautiful road is a matter of preferences and, as such, non-rational in nature, the irrational choice would be the one in which one ended up at point C instead of B, insofar as this implies that something has gone wrong in the process.

Inspired by Hobbes, one might also argue that an action is instrumentally irrational if it undermines the primary aim of rationality, self-preservation, when it turns from a matter of reaching one's desires to a blind pursuit regardless of the costs to oneself, and socially irrational when this blind pursuit of desires happens at the expense of others and turns the social world into a state of war. Irrationality may, then, also be when the agent acts against her own self-preservation. Plato would argue, in contrast to this consequentialistic perspective, that an irrational action is one which is based on a wrong set of goals, beliefs, etc. and, as such, the action would be opposite to reason and would invoke an internal war between the three masters of the soul. In a similar manner, Kant would argue that an action is irrational because it is based on an irrational foundation –

a will which is not in command of itself through reason, a will which allows itself to be influenced and in that manner subdue itself to lower passions or the will of others - letting itself become a means to someone's end. Irrationality related with self-governance cannot, as Plato seems to argue, be reduced to a matter of mental faculties, insofar as the will's control over itself is not solely contingent upon matters related with epistemology, but also matters related with social causality. From the point of view advocated by Smith, one might argue that it is irrational to believe that one is not influenced by other individuals and their opinions, because this sensitiveness not just influences one's self-perception, it is also the very foundation of morality – being sensitive to other agents is the very foundation of social rationality, in the sense that it allows the agent to transcend her own private interests. Indeed, if life does not unfold itself in a social vacuum it is irrational to believe that one is not influenced by other social agents, just as it is irrational to believe that one is entirely in control of one's own actions, insofar as social life implies an exchange of rights and claims. But that does not liberate the agent from the responsibility which accompanies her actions; although she might not be entirely in control of actions, these are still produced and as a producer of actions, she is responsible to herself as well as to others. In that sense, it is socially irrational to believe that one is ever without responsibility. The relation between causality and blame-ability will be discussed in chapter VI.

In sum, then, it might be argued that irrationality is related with the causal process of production and, as such, it is purely non-instrumental, because it is not related with the consequences of one's actions. Irrationality is related with the cognitive foundation of the action and not the action itself, thus irrationality is related with the relationship between action and judgments, just as it is related with the reasons one have to endorse certain beliefs and values. Irrationality does not, however, imply a lack of reason, but it might imply wrong reasons. Matters of choice are matters of preference and preferences cannot be irrational or rational. It is not irrational to act morally if acting otherwise brings about consequences which are against one's self-preservation – then it is socially rational to act in accordance with the moral code. Rationality can, then, not be reduced to a binary distinction between actions which has oneself as benefactor and actions which has the other as benefactor, insofar as rationality is linked with morality through the notion of social rationality associated with social interaction – rationality then co-

vers how one governs oneself, regain control of oneself, but it also covers how one governs oneself in the company of others, and how one interacts with others. Morality is about how individuals should act in relation to other social agents and, as such, rationality without morality is asocial, just as morality without rationality is unrealistic. What matters is not whether individuals are entirely egoistic or altruistic, rather what matters is how it is balanced. Social rationality takes this balance seriously and as a concept it is aimed at capturing the middle ground between rationality and morality. This meta-perspective is also accordance with the utilitarian tradition discussed, just as it is in accordance with the Plato-Kantian perspective. In that sense, an action might be against one's private interests, but in accordance with what is perceived as socially rational, just as the opposite might also be the case. An overall rational action is, then, both in accordance one's private interests and the interests of other social agents – hence, the action must be morally as well as instrumentally justifiable. Now, an action may be justifiable in terms of one's goals, just as the means one apply in the pursuit of them may be logical, however, the values represented in the goal may not be socially justifiable. To some extent, then, this allows one to understand the cognitive foundation of Folger's (2012) argument that some goals were justifiable, but the means necessary to bring them about may not be legitimate, in the sense that justification, here, is related with individual rationality and legitimacy with social rationality. Rationality cannot, then, be opposite to morality – without presupposing asocial behavior.

Rationality and social rationality cannot, however, be viewed as producing infallible and universal decisions, insofar as they are based on the agent's limited abilities. Rationality is subjective and to act rationally means nothing but to act on non-defeated reasons. Besides being subjective, rationality is also normative in nature, in the sense that rationality consists in the intelligent pursuit of justifiable ends – as such, it is related with the employment of justifiable and non-defeated reasons, a social agent is rational if her actions are governed by these strong reasons and if these reasons generate the foundation of her judgments on matters related with action, evaluation, and cognition. Rationality is, then, about what the agent *should* believe, do, and value; and the rational agent is one who allows her judgment to be based on and governed by the strongest reasons.

In accordance with what has been argued so far, it seems to follow that rationality cannot be separated from fairness, insofar as rationality is related with the very process of producing these judgments. Of course, rationality, as described here, does entail neither egoism nor omnipotence and, as such, it deviates from the models which fairness theories are generated in opposition to. However, Folger (2012) might also be correct when he argues that the non-morally founded theories on fairness simply substitutes a focus on monetary outcomes with a focus on non-monetary outcomes, either in terms of control or socioemotional experience. Although, even a morally founded theory on fairness cannot be separated from its rational foundation, at least not without losing its intellectual roots or producing an unrealistic conception of social agents as purely concerned with the well-being of others.

From a horizontal perspective, then, fairness is related with the production of rational and socially rational actions. As such, it is related with the production of actions and the values which these come to represent along with the goals which one aims at bringing about, just as it is a matter of understanding that one's actions has implications for other social agents and that these implications might have consequences for oneself or, at least, for the relation between the acting and the affected agent. It should not be perceived as an argument against selfish behavior – against behavior which has as its aim to bring about some state in which the acting agent is the only benefactor – rather it should be perceived as an argument against selfishness being the only social value, insofar as if this was the case, the social system would collapse into a state of war. Hence, social rationality has as its aim to understand how selfish behavior is balanced with non-selfish behavior. Besides being a felt obligation (in accordance with deonance), social rationality can also be viewed as a socially produced obligation – a pragmatic necessity – which follows as a consequence from the fact that taking part in organizational life implies taking part in social life and, as such, it entails that the social agent is influenced by her colleagues. That is, her opportunity set is not just influenced by the rights she has passed vertically, but also by the rights she has passed horizontally by taking into consideration how she should act towards them. In that sense, this cannot be reduced to extra-ordinary considerations related with abnormal (non-normal) actions; rather it is related with how social agents in their everyday life balance their private interests with the interests of others. Additionally, it cannot be reduced to a matter of identifying the

distribution which is considered as being the fairest one or the principle which is considered as the fairest one. Rather, it becomes a matter of understanding causality – the act of bringing some particular state about, the means which goes into this, and the actual state. That is, the agent must ask herself if the state she wants brought about is justifiable, whether or not the means are justifiable, and whether or not she actually wills the end and the means – whether or not these are in accordance with her own principles – what she considers as good, as right, as fair, etc. Although, the answer is subjective, the agent must be able to justify it to others and, in that sense; she must not be able to produce any strong reasons against doing it.

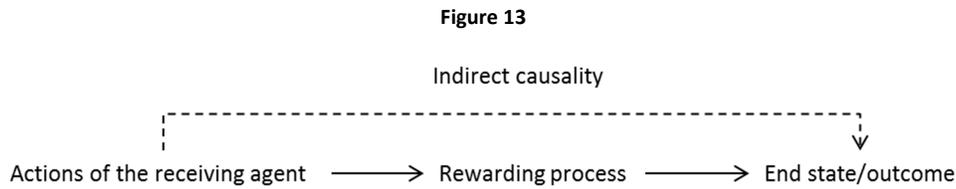
## Chapter IV

### Fairness as Reciprocity and Non-Economic Motives

#### 4.1 Introduction

In the previous chapter, social rationality was identified as the foundation as the basis of a horizontal conceptualization of fairness, just as the philosophical foundation of the three metatheoretical approaches to vertical fairness theories was discussed. Social rationality was defined as the normative judgment on how the agent should act and the values she should endorse. Moreover, social rationality was defined as the social dimension of rationality and rationality was defined as acting, believing, and evaluating on strong reasons. Social rationality as well as rationality was defined as being subjective and normative in nature. Additionally, it was also argued that horizontal fairness, in opposition to vertical fairness, could not solely be understood as fairness judgments related outcomes, insofar as the focus here is on the causal production of different states – a perspective which enables one to take into consideration the fact end states as well as the actions which brings them about can be viewed as transgressions of foundational relational norms inherent to the horizontal relation. That is, if horizontal fairness is to push past the notion of fairness developed in equity theory (see the next chapter) it has to take into consideration the causal production of end states and the fact that unfair behavior or antisocial behavior may also emerge as matters related with the behavior espoused by the receiving agent – the agent who, for example, differentiate herself by supplying a higher level of effort. What matters, in other words, is not just the outcome difference, but also how it was produced. Causality is, of course, a problematic concept in relation with horizontal fairness, insofar as the agent cannot, directly, be blamed for bringing some state about and, as such, she may not be responsible for bringing about the end state, but only for producing the actions which indirectly led to the end state. I return to the matter of causality in chapter VI, however, a tentative definition of indirect causality might be helpful: the effects of indirect causality are nested in the total result and can only be approached through counterfactual language related with what would have happened had something not occurred – that is, the indirect effect X upon Y might

be analyzed by holding X constant and change the mediating variable Z (Pearl, 2009: 165). This might be illustrated as follows:



The causal relation between the actions of the receiving agent and the reward process was discussed in chapter II, while the causal relation between rewarding and different end states and their relations with vertical fairness was discussed in part I of chapter III. As the figure illustrates, a direct relation between the rewarding process and the fairness of this seems to exist, whereas, from a vertical relation, the indirect relation between the actions of the receiving agent and the end state can solely be indirectly approached through counterfactual reasoning about which state should have occurred had the acting agent not acted the way she did – had she acted, for example, in accordance with governing norms, see chapter VI.

Theories based on different distribution games (for example the dictator game) may provide some insight into the behavior of the receiver and the distributor, just as it might provide some insight on how the receiver evaluates the offer, when she accepts, etc. The perspective, however, cannot provide insight on vertical fairness, insofar as one cannot presuppose direct causality between the actions produced by the receiving agent and the end states. That is to say, the receiving agent – the agent receiving the reward – is not directly responsible for bringing about a particular outcome difference as the distributor is in distribution games is assumed to be. In addition, as the two examples at the end of chapter II explicated, the fairness of the end state, from a vertical perspective, cannot be reduced to matters related solely with outcome differences, but must also take into consideration the actions which produced this end state.

In the previous chapter, I aimed at discussing the normative foundation of horizontal relations. The normative foundation, however, cannot be reduced to horizontal relations, insofar as the employment relation is also normatively centered upon matters related with vertical relations, the relation between the employee and the organization. In

that sense, in order to produce a holistic conceptualization of the normative relational foundation of organizational life one also needs to take into consideration how these are structured, formally and as well as socially. Thus, the aim of this chapter is to understand the normative foundation of vertical relations.

#### **4.2 Economic Rationality and Fairness as Reciprocity**

Recent models of the normative foundation of vertical relations have focused on reciprocity. Reciprocity, as a model on the normative foundation of fairness has its intellectual roots in Aristotle's (350BC/2007) four basic kinds or modes of fairness described in his work "Ethics". Social rationality, as described in the previous chapter, can be resembled with the kind of fairness, Aristotle defined as reasonableness – the highest kind of fairness which emerge not from force or centered upon any criteria besides the one which the individual commits herself to – reasonableness, then, is self-commitment, in the sense that the agent commits herself, binds herself, to treat her fellow humans in a particular manner. In this chapter, I will discuss reciprocity, in chapter V the focus is on equity (fairness as proportionality) the most basic kind of fairness identified by Aristotle focusing on the relational dimension of fairness, in chapter VI the focus shifts to a focus on the cognitive dimension of fairness related with unfairness as abnormal events (corrective fairness) – deviations from what is normal – and in chapter VII the temporal dimension of Aristotelian fairness will be discussed. The Aristotelian fairness model is here used to reconstruct not just the different approaches or modes of fairness, but also in order to reconstruct the different dimensions or circumstances which go into the construction of the *Sachverhalt* or, translated into English, the issue or the case at hand.

Reciprocity to Aristotle (350BC/2007: V.5, 1132b/131) is retribution, in the sense that the actions or goods exchanged are equivalent in kind or value. Reciprocity, nevertheless, is neither proportional fairness nor corrective fairness, in the sense that it does not require equality and therefore not the restoring of it (Aristotle, 350BC/2007: V.5, 1132b/132). Reciprocity concerns the proportionality of daily interactions on which a community exists, that is, social exchanges between social agents. It requires, in other words, three important matters: a) two agents with opposite demands (e.g. a buyer and a seller); b) the proportional relation of the goods/burdens/actions which are to be ex-

changed must be established, e.g. by the use of a mediating system of measurement making the goods comparable – such as money; and c) the fair exchange requires that the exchanging agents’ situations, respectively, after the exchange has occurred, is equal to their situation before the exchange – that is, none of the parties must be made worse off than before the exchange occurred (Aristotle, 350BC/2007: V.5, 1133a/132, 1133b/133).

Reciprocity, as it is discussed by Aristotle (350BC/2007), captures the notion of reciprocity as the exchange of opposites equivalent in kind or value, as such, the exchange will have an equalizing effect on the social relation in regard to the underlying temporal demarcation of the exchange into “a before” and “an after”. Combined the result created by the exchange must be equal to the combined sum of the two parties’ individual contributions – if not it deviates from *Pythagorean* reciprocity<sup>27</sup>. This implies that the actions must be scaled to one another in terms of kind and effect. The actions of the reciprocal agent, then, are not centered on material future benefits, rather the actions are responses to the other social agent’s actions regardless of possible material gains or costs (Fehr and Gächter, 2000: 161).

Matthew Rabin’s (1993) model in psychological game theory which seeks to capture the effect of reciprocity in dyadic relations might be referred to as one of the foundational models of the economics of fairness. Rabin (1993: 1282, 1284) based his idea on three basic assumptions: (1) people are willing to sacrifice private well-being to reciprocate kindness; (2) people are willing to sacrifice private well-being to reciprocate unkindness; and (3) as the costs of reciprocating increase the reciprocating individual’s motivation to reciprocate decreases. Suppose  $S_1$  and  $S_2$  are two strategies that the players 1 and 2 can choose, and let the material payoff of Player  $i$  be represented by the following function,  $\pi_i: S_1 \times S_2 \rightarrow \mathbb{R}$  (Rabin, 1993: 1286). Furthermore, Rabin (1993: 1284) argues that the players’ payoff depend simultaneously on their actions and their beliefs about the other players’ actions. Player 1 chooses an action,  $a_1$ , when she believes that Player 2 has chosen,  $b_2$ . Rabin (1993: 1286), now, develops a “kindness function”

---

<sup>27</sup> Pythagorean reciprocity refers to Pythagoras’ theorem on the geometry of a right triangle – the square of the hypotenuse is equal to the combined product of the two other sides squared.

measuring how kind Player 1 is to Player 2 when she chooses  $a_1$  while believing that that Player 2 chooses  $b_2$ :  $f_1(a_1, b_2)$ .

Let  $\pi_2^h(b_2)$  be Player 2's highest payoff in the game. Now, let  $\pi_2^l(b_2)$  be Player 2's lowest Pareto efficient payoff in the game. The equitable payoff can now be defined:

$$\pi_2^e(b_2) = \frac{\pi_2^h(b_2) + \pi_2^l(b_2)}{2}$$

Now, let  $\pi_2^{min}(b_2)$  be the worst possible outcome for Player 2. The kindness of Player 1 to Player 2 can now be defined:

$$f_1(a_1, b_2) \equiv \frac{\pi_2(a_1, b_2) - \pi_2^e(b_2)}{\pi_2^h(b_2) - \pi_2^{min}(b_2)} \text{ if } \pi_2^h(b_2) - \pi_2^{min}(b_2) = 0, \text{ then, } f_1(a_1, b_2) = 0$$

This represents how kindly Player 1 believes she is treating Player 2, when playing  $a_1^1$ . Player 1 is kind to Player 2 if she gives him more than the equitable split – that is, if  $\pi_2(a_1, b_2) > \pi_2^e(b_2)$ . The degree of kindness is scaled by the possible payoffs Player 2 could have received. Player 1, of course, is only kind to Player 2 if she believes he is kind to her – this Rabin (1993: 1287) captures in the following function:

$$\tilde{f}_2(c_1, b_2) \equiv \frac{\pi_1(c_1, b_2) - \pi_1^e(c_1)}{\pi_1^h(c_1) - \pi_1^{min}(c_1)} \text{ if } \pi_1^h(c_1) - \pi_1^{min}(c_1) = 0, \text{ then, } \tilde{f}_2(c_1, b_2) = 0$$

Where  $c_1$  represent Player 1's beliefs about Player 2's beliefs about Player 1's actions. This function, then, captures how fair Player 1 judges her action to be given her beliefs about Player 2's beliefs. The expected utility function can now be written:

$$U_1(a_1, b_2, c_1) \equiv \pi_1(a_1, b_2) + \tilde{f}_2(c_1, b_2)(1 + f_1(a_1, b_2))$$

Player 1 choose  $a_1$  which affects her own kindness function  $f_1(a_1, b_2)$ . If, on the one hand, Player 1 believes Player 2 to be kind, then,  $\tilde{f}_2(c_1, b_2) > 0$  – thus her utility is increasing as a function of her own kindness. On the other hand, if Player 1 believes Player 2 to be unkind, then her utility is decreasing as a function of her own kindness - im-

plying that she obtains positive utility from hurting the other player.<sup>28</sup> Consider now, a standard Battle of the Sexes:

Figure 14: Rabin's Battle of the Sexes Game (Rabin: 1993: 1285).

		Player 2	
		Opera	Boxing
Player 1	Opera	X,2X	0,0
	Boxing	0,0	2X,X

Now, Player 1 dislikes going to the opera, but prefers going to boxing. Player 2 dislikes going to boxing, but prefers going to the opera. Both players, however, prefer to spend the evening together. The game has mixed Nash equilibria, because Player 1's best response if Player 2 plays opera is opera, and if Player 1 plays opera, then Player 2's best response is also opera. The same applies for boxing. Are the Nash equilibria also a fairness equilibrium under Rabin's model? To answer this question, one needs to take into account the two agents' first and second order beliefs – that is, beliefs about the other player's actions and beliefs about the other player's beliefs (fairness). Hence, if Player 1 believes that Player 2 is going to play opera, and believed that Player 2 believed that Player 1 would play opera – based on this, would Player 1 prefer to play opera or boxing? The same line of reasoning goes for Player 2.

$$\tilde{f}_2(O, O) = \frac{\pi_1(O, O) - \pi_1^e(O)}{\pi_1^h(O) - \pi_2^{min}(O)} = \frac{2X - 2X}{2X - 0} = \frac{0}{2X} = 0, \text{ hence:}$$

$$U_1(a_1, b_2, c_1) = \pi_1(a_1, b_2)$$

Thus, Player 1 does not believe that Player 2 is being neither fair nor unfair, so Player 1 prefers to play *opera*. The same applies for Player 2, hence {O,O} is an equilibrium. Consider now if Player 1 believes that Player 2 is plying *boxing*, while also believing that Player 2 believes that Player 1 plays *opera*. Would Player 1 prefer playing *opera* or *boxing*?

<sup>28</sup> Note  $1 + f_1(a_1, b_2)$  which entails that whenever Player 2 is unkind to Player 1, then Player 1's payoff her material payoff.

$$f_1(O, B) = \frac{\pi_2(O, B) - \pi_2^e(B)}{\pi_2^h(B) - \pi_2^{min}(B)} = \frac{0 - 2X}{2X - 0} = -1$$

$$\tilde{f}_2(O, B) = \frac{\pi_1(O, B) - \pi_1^e(O)}{\pi_1^h(O) - \pi_1^{min}(O)} = \frac{0 - 2X}{2X - 0} = \frac{-2X}{2X} = -1$$

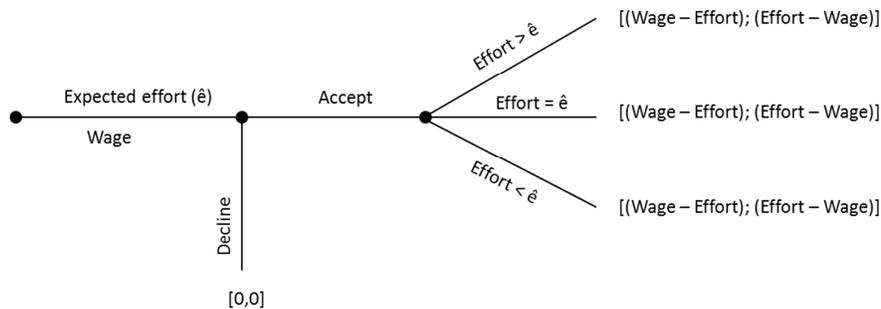
$U_1(O, B, O) = 0 + (-1)(-1 + 1) = 0$ , while if playing *boxing*:

$$U_1(B, B, O) = X - (1 - 0) = x - 1$$

This shows that if  $x$  is small enough, then, Player 1 is prepared to play *opera* to spite Player 2 if she believes that Player 2 is being unfair. {B,O} and {O,B} are spiteful equilibria. According to Rabin (1993: 1283), fairness matters most when payoffs are small, thus if the payoffs are sufficiently large, then, fairness plays a lesser role.

Reciprocity is, then, related with the exchange of actions equivalent in kind and effect as argued by Aristotle. However, reciprocity is also the attribution of intentions, insofar as Rabin's (1993) model seems to demonstrate that the alternative action which the agent could have chosen becomes a signal of intent. In chapter VI, I will discuss the cognitive normative foundation of intentionality. From this it seems to follow that reciprocity is of first order importance, because it changes the underlying structure of the game and, as such, it might induce selfish players to take non-selfish actions if they expect that the other part will punish them – hence, it alters the first-order preference structure of the game (Fehr and Fischbacher, 2002: C4). This entails that reciprocity may reduce the efficiency of explicit incentive programs designed to enhance efficiency, insofar as it creates implicit incentives (Fehr and Fischbacher, 2002: C22, C28). Consider the game below:

Figure 15: Illustration of the game on voluntary cooperation



At the first decision node, the employer chooses a wage and an expected effort level which she offers the potential employees. The expected effort level is not binding for the employee. The employee, now, chooses whether he accepts or declines the offered contract, if he declines, the game ends. On the other hand, if the potential worker accepts the contract, then he has to choose an effort level which can either be below, above or equal to the expected effort level. According to agent-theory the employee should have no incentive to choose an effort above minimum, just as the employer believing the employee will choose  $e^{\min}$  will have no incentive to set the wage above  $w^{\min}$ . What the authors find, however, is that the higher the rent (wage –  $\hat{e}$ ) offered, the higher actual effort levels was chosen by the employee. This implies, in accordance with the rule of reciprocity that kind gestures are responded to in a kind manner (Fehr and Falk, 2002: 691). Fehr and Gächter (2000) argue on this matter:

“The requirement of a generally cooperative job attitude renders reciprocal motivations potentially very important in the labor process. If a substantial fraction of the workforce is motivated by reciprocity considerations, employers can affect the degree of “cooperativeness” of workers by varying the generosity of the compensation package – even without offering explicit performance incentives.” (Fehr and Gächter, 2000: 171)

This implies two important matters: (1) in accordance with Truman Bewley’s (1998: 475; 2004: 6) theory on downward wage rigidity as being caused by concerns about employee morale and that the employment relation is based on certain norms concerning one’s own behavior which also generate expectations on how other social agents should act; and (2) voluntary cooperation between the employee and the organization is more important than cooperation emerging from coercion or other kinds of extrinsic influence. This matter leads the authors to their idea on *belief management* centered on the idea of reciprocal agency: Fehr and Falk (2002) argue on this point:

“The existence of conditional cooperation renders the management of the workers’ beliefs about other workers’ effort important because if a conditional cooperator believes that the others shirk he will also tend to shirk.” (Fehr and Falk, 2002: 692)

Belief management, here, concerns the employees’ beliefs on fairness not solely in regard to the employment relation (compensation), but also in relation to the other employees – on the latter the authors argue that management has an important role, because direct management should be used to impute fairness (equitability) in the em-

ployment relation by removing non-cooperative workers, just as they should be concerned with hiring the right people. Belief management, in other words, is concerned with sending the proper signals in relation to both compensation and effort, insofar as fairness is a socially constructed belief which emerges from comparison of one's own situation with that of similar others (a point which was also mentioned by Aristotle, see the previous chapter). Reciprocal fairness, then, is not solely contingent on the behavior of the parties directly involved in the exchange, but also on the behavior of others, a third party, insofar as their exchange relation influences how the agent feels about her own exchange. That is, the voluntary cooperation of the employee depends on it being appreciated and reciprocated by the management, just as it depends upon the behavior of the other employees, thus fairness in the employment relation depends upon the vertical relation between management and the employee, and the horizontal comparison of one's colleagues' vertical relations, especially in relation to judging the fairness of one's contribution.

The authors identify, in other words, that loyalty is contingent upon the behavior of other individuals – implying that the social agent is only conditional loyal. This is different from how loyalty has been discussed by, among others, Herbert Simon (1991: 35, 37, 38) who argues that selfish motivation cannot explain employment relations, rather the employment relation is centered upon three psychological phenomena: *docility*, *identification*, and *bounded rationality*. Simon (1983: 65; 1991: 35) bases his idea about docility on an idea about human nature which through evolution has been made cooperative or civilized to act in a manner which is socially approvable. Identification, actually first introduced in organizational theory by Katz and Kahn (1966/1978: 374), relates to the psychological necessity of belonging to a group – in way to be part of a “we” as opposed to a “they” (Simon, 1991: 36). Lastly, because social agents are bounded rational they do not possess the mental capacity to make decisions which fully takes account of all factors, thus one way of reducing complexity is through adopting the goals of one's organization or department – so, by attending to these goals they are contributing to the “we” (Simon, 1991: 37). More commonly this would probably be referred to as displays of organizational citizenship behavior (see, for example, Dennis Organ, 1997, for a theoretical review).

The study referred to above may be viewed as a classic gift-giving game originally introduced by George Akerlof (1982). Puzzled by the results of one of Homans' (1954) studies "The Cash Posters", he studied, why the observed actual effort level was above the minimum installed by management. From an economic perspective, the employees should have no incentive to supply a higher level of effort, because they were not rewarded and there were no consequences from not obtaining the minimum rate, just as the job was not perceived as a career. Furthermore, no social norms could be identified which could explain this. Akerlof (1982: 544) then tried to explain this by introducing the notion of gift-giving inspired by Marcel Mauss's (1922/1966) idea on the logic of gift-giving in archaic societies. Mauss (1922/1966) defines the obligation which accompanies the gift in the following way:

"The obligation attached to a gift itself is not inert. Even when abandoned by the giver, it still forms a part of him. Through it he has a hold over the recipient..." (Mauss, 1922/1966: 9)

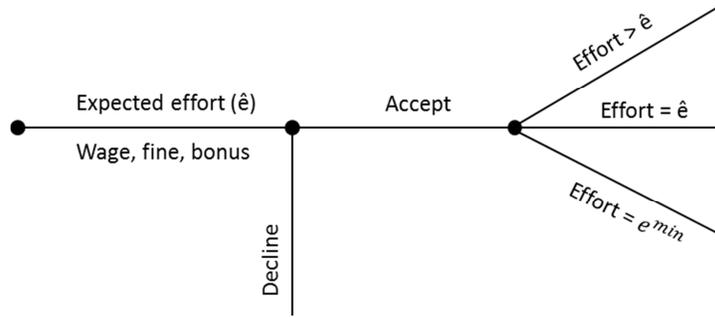
He goes on arguing that:

"... a return will give its donor authority and power over the original, who now becomes the latest recipient. That seems to be the motivating force behind the obligatory circulation of wealth, tribute, and gifts..." (Mauss, 1922/1966: 10)

In accordance with this, Akerlof (1982:544) argued that the excess effort (actual effort > effort minimum) could be viewed as the employees' gift to the organization, while the organization gave the employees a wage which was above the competitive level. This notion led Akerlof (1982: 546, 547) to criticize neoclassical contract theory for neglecting these social dynamics and presupposing egoistic behavior.

This particular understanding of the employment relation also has implications for the use of performance pay, because the underlying assumptions on self-interested actions driven by economic motives may clash with the underlying logic of reciprocal relations. Fehr and Gächter (2002) and Gächter, Kessler, and Königstein (2008) extended Fehr, Gächter, and Kirchsteiger's (1997) study, mentioned above, by adding the possibility of rewarding and sanctioning. The game can be illustrated as follows:

Figure 16: Illustration of the extended gift-giving game.



The introduction of the possibility of imposing sanctions and rewards changed the nature of the game. In the game without incentives, the trust game, the employer initially determines the contract within the ranges:  $1 \leq \hat{e} \leq 20$  and  $-700 \leq w \leq 700$  (Gächter, Kessler, and Königstein. 2008: 3). If the employee accepts the offer the following payoffs will emerge:

$$\pi^{Employee} \left\{ \begin{array}{l} \text{Accept} = w - (7e - 7) \\ \text{Decline} = 0 \end{array} \right\}$$

$$\pi^{Employer} \left\{ \begin{array}{l} \text{Accept} = 35e - w \\ \text{Decline} = 0 \end{array} \right\}$$

Because the wage does not change as a function of effort, the employee has no incentive to choose an effort above  $e^{\min} = 1$  where  $c(e) = 7(1) - 7 = 0$  implying that the employee minimizes her costs of effort. Anticipating this, the employer will have no incentive to choose a  $w$  higher than 1 (Gächter, Kessler, and Königstein. 2008: 3). In the best of all worlds, the employee would choose  $e = e^{\max} = 20$  because it maximizes the total surplus:  $35e - (7e - 7) = 700 - 133 = 567$ .

If a negative incentive is introduced, the employer has to choose  $w$ ,  $\hat{e}$ , and  $f$ . The fine ( $f$ ) is subtracted from the employee's wage and added to the employer's profit, thus the following payoffs now emerge (Gächter, Kessler, and Königstein. 2008: 4):

$$\pi^{Employee} = \left\{ \begin{array}{l} \text{Accept}; e \geq \hat{e} = w - (7e - 7) \\ \text{Accept}; e < \hat{e} = w - (7e - 7) - f \\ \text{Decline} = 0 \end{array} \right\}$$

$$\pi^{Employer} = \begin{cases} Accept; e \geq \hat{e} = 35e - w \\ Accept; e < \hat{e} = 35e + f - w \\ Decline = 0 \end{cases}$$

The employee will, now, have no incentive to choose an effort level different from  $e = e^{\min}$  or  $e = \hat{e}$ . The employee should choose  $e = \hat{e}$  if  $w - c(\hat{e}) \geq w - f - c(1) \rightarrow f \geq c(\hat{e})$ . The employee should, then, choose  $e = \hat{e}$  if the fine is greater than or equal to the costs of supplying the necessary level of effort – otherwise she should choose  $e = e^{\min}$ . This also produces a constraint on the contractual design, insofar as accompanying each level of a fine is a maximal effort which can be enforced – for example, if the fine is 24 the maximal effort which this may produce is:  $24 \geq 7e - 7 \rightarrow -7e \geq -7 - 24 \rightarrow e \leq 4.4$ . A self-interested employer will maximize effort and choose the highest fine possible, here 80, which makes the highest level of enforceable effort,  $e = 12$ ,  $w = 7(12) - 7 = 77$ , the employer's gross profit =  $35(12) = 420$ , and the total surplus = 343. The employee's best response here, then, is to choose  $e = \hat{e}$ .

In the positively framed incentives game the employer states  $[w, \hat{e}, b]$ . The bonus ( $b$ ) is added to the employee's wage and subtracted from the employer's earning, it is only given if:  $e \geq \hat{e}$ . The following payoff structure now emerges (Gächter, Kessler, and Königstein. 2008: 6):

$$\pi^{Employee} = \begin{cases} Accept; e \geq \hat{e} = w + b - (7e - 7) \\ Accept; e < \hat{e} = w - (7e - 7) \\ Decline = 0 \end{cases}$$

$$\pi^{Employer} = \begin{cases} Accept; e \geq \hat{e} = 35e - w - b \\ Accept; e < \hat{e} = 35e - w \\ Decline = 0 \end{cases}$$

The game is centered on a similar reasoning as the fine game, above, because the employee will have no incentive to choose  $e > \hat{e}$ , because then the positive effect of the bonus would be lost which implies that  $e^{\min}$  is optimal if  $e$  different from  $\hat{e}$ . Hence, the employee will choose  $e = \hat{e}$  if:  $w + b - c(\hat{e}) \geq w - c(1) \rightarrow b \geq c(\hat{e})$ . In a similar way as with a fine, a bonus will also produce a limited effort effect – a constraint stated by inequality. Like with the negative incentive, the employer should choose the highest bonus possible, here 80,  $e = 12$ ,  $w = [7(12) - 7] - 80 = -3$ . The self-interested employee

will accept the contract and her best response will be  $e = \hat{e}$  (Gächter, Kessler, and Königstein. 2008: 6).

The games described above are based on the assumptions in neoclassic contract theory, which states that the incentive contract is more efficient than the trust contract. In the trust contract, voluntary cooperation (or the gift given to the organization) is  $e - e^{\min}$ . The games described, here, have three phases in which the respondents shifted between the different treatments, described above, just as they changed the order of the treatments to study the effects of, for example, changing from a trust game to a fine game (Gächter, Kessler, and Königstein. 2008: 7). If the respondents were exposed to a pure trust game meaning that all three phases was trust games, the authors identify a positive correlation between wages offered and the effort chosen by the employees – that is, the more generous the rent offered ( $\hat{e}-w$ ) the higher effort level supplied (Gächter, Kessler, and Königstein. 2008: 11). Across all cases and phases, the effort level chosen was above the employees' best response predicted by neoclassical contract theory,  $e = e^{\min}$ . By allowing the employers to state a fine or a bonus, the authors could also study the possible interaction between reciprocity and incentives – the social crowding-out effect. Compared to the pure trust game, the incentive games were less efficient, because across almost all cases, the effort level was below the enforceable effort level ( $e = 12$ ), that is, the employees' did cooperate voluntarily. Additionally, the introduction of incentives did not change the average effort level when comparing the incentives games to the pure trust games – hence, the introduction of incentives crowds-out the desire to voluntary cooperation, because incentives substitute trust (Gächter, Kessler, and Königstein. 2008: 14). This crowding-out effect is also present in the long run, that is, if the employees' shifts from an incentives game in the first phase to a trust game in the second phase – the effort levels chosen in the second level are below the ones chosen in the pure trust game, the crowding-out effect of fines, however, is higher than the one of bonus (Gächter, Kessler, and Königstein. 2008: 16).

Combined the studies show how reciprocity interacts with cognitive framing procedures, because if the incentive is framed in a positive manner as a reward, the effort level is higher than if the incentive was framed in a negative manner as a sanction, albeit in both cases the level of effort was below that in the baseline without incentives. The

reason is, the authors argue, that the interpretation of an act as hostile/kind is contingent upon a frame of reference – in the negative framing, the frame of reference is total compensation (wage + bonus), thus being caught shirking implies that one loses out on the bonus and in a sense this is equivalent to the bonus being taken away or subtracted; whereas, on the other hand, a positive framing as a bonus implies adding because here the frame of reference is the base wage and as such something is given to the employee. The basic argument put forth, here, is that the introduction of incentives out-crowds pro-social behavior such as displays of loyalty, because it changes the underlying structure of the game from one being based on, for example, reciprocity or some other social relational norm to non-social relation regulated solely through incentives.

The possible substitution of pro-social fairness concerns with economic concerns was also demonstrated by Uri Gneezy and Aldo Rustichini (2000) who studied the interaction between fairness concerns and economic concerns in a kindergarden in the relation between parents and the employees working in the kindergarden. Initially, the parents who were late to pick up their children felt guilty, because their behavior had consequences for the employees working in the kindergarden and as a consequence of this, they would be on time another day – that is, by being late the parents violated the social contract existing between the employees in the kindergarden and themselves. To overcome this problem, the kindergarden imposed a fine on being late (Gneezy and Rustichini, 2000: 4). The imposition of the fine, however, had an adverse effect, because what was previously governed by fairness concerns was now governed by economic concerns; hence the parents just arrived late and paid the fine – the authors even found that the number of parents arriving late increased during the period with the fine (Gneezy and Rustichini, 2000: 8). As also showed in the example above, the introduction of the fine had long term consequences for the relation between the employees and the parents, because even after the fine was removed, the parents still arrived late. In that sense, if economic norms override social norms, the possible negative effects will be long term (Gneezy and Rustichini, 2000: 8).

In another study by Fehr, Alexander Klein, and Schmidt (2007) the authors find that in terms of efficiency, the bonus contract outperforms the negative incentives contract and the pure trust contract. The employers were, now, to choose between offering a

negative incentives contracts or a bonus contract, and between offering a trust contract and offering a negative incentives contract. Neoclassical contract theory predicts that both the bonus and the trust contract will be outperformed in terms of efficiency compared with the negative incentives contract – hence, the authors’ results contradicts this assumption, at least, in regard to the relation between the negative and positive incentives contract (Fehr, Klein, and Schmidt, 2007: 140). In the choice between offering a trust contract and offering a negative incentives contract, the employers in the first round seemed almost indifferent between the two, however at the beginning of the fourth round almost 80 percent of the employers had shifted to negative incentives contracts (Fehr, Klein, and Schmidt, 2007: 129). The average chosen effort across all the periods for the pure trust contract was 1.98 which is only slightly higher than  $e = e^{\min} = 1$ . In this study, the respondents could choose effort within the range [1-10] implying that the efficient effort level in the best of all worlds is 10 which yields a total surplus of:  $10(10) - 20 = 80$  (Fehr, Klein, and Schmidt, 2007: 127). The negative incentives contract which the employers shifted to because they learned that the pure trust contracts performed badly were in most cases non-compatible<sup>29</sup> incentive contracts, because the wage offered to the employees was too high or the expected effort was too high (maximum fine was 13 which yields that the highest enforceable effort is 4) – the generous high wage did not invoke reciprocal behavior, while the high expected level made the employees shirk, and both led to negative payoffs to the employers (Fehr, Klein, and Schmidt, 2007: 131, 134). If these results are viewed in relation to Fehr and Schmidt’s (1999) model, see chapter V, it seems that the proportion of reciprocal minded employees was too small to make the employers prefer the pure trust contract or the non-compatible incentives contract. This model also explains why the employers preferred the incentives contract to the pure trust contract – intuitively the incentive contract had a higher enforceable effort level than the trust contract. Additionally, if the proportion of reciprocal minded employees is too small, then the marginal effort effect produced by a marginal increase of the employees’ wage is too small and leads, eventu-

---

<sup>29</sup> An incentive-compatible contract is a contract designed to ensure a mutually beneficial behavior by both parties, thus a non-compatible incentive contract, here, is one which is not in accordance with either the assumption on wage stating that the employee should only obtain  $w^{\min}$  providing her with her reservation utility (0), or the assumption that the negative incentive effect is only valid within a certain of effort – that is, the enforceable level of effort.

ally, to a decrease of the employers' profit (Fehr, Klein, and Schmidt, 2007: 146). If the proportion of reciprocal minded employees is 0.4, then:

$$\begin{aligned} \text{Maximization of payoff: } \pi^{\text{Employer}} = \pi^{\text{Employee}} &\rightarrow 10e - w = w - c(e) \\ &\rightarrow 10e + c(e) = 2w. \text{ Differentiation then yields } \frac{\Delta e}{\Delta w} = \frac{2}{10 + c'(e)} \end{aligned}$$

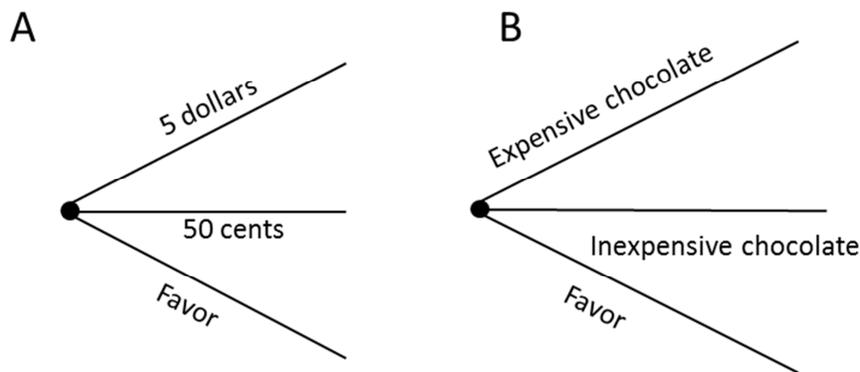
A marginal wage increase yields:  $0.4(2/[10+1]) = 0.07$  and the employer's gross profit increases by at most:  $10 \times 0.07 = 0.7$ , hence a wage increase is associated with a loss of profit (Fehr, Klein, and Schmidt, 2007: 145). The incentives contract, nevertheless, might be rejected by the inequality-averse employee because the profits obtained by the parties are unequal: Employee profit =  $4 - 4 = 0$  and Employer profit =  $10(4) - 4 = 36$ . Inequality aversion, then, might lead the employer to propose and prefer the incentives contract to the pure trust contract, just as inequality aversion might lead to the potential employee to reject the offer.

In the second case when the employers could choose between offering a non-binding bonus contract and a negative incentives contract, the majority of the employers chose the bonus contract (Fehr, Klein, and Schmidt, 2007: 135). The average level of effort for the bonus contract was 5.22 which were above the highest enforceable effort level (4 when the maximal bonus is 13). The bonus contract also produced higher pay-offs to both parties, and the contracts offered was designed in way which produced low base pay and high bonuses (Fehr, Klein, and Schmidt, 2007: 137). In that sense, the introduction of a possible bonus invoked fairness concerns within the employees based on the cognitive structuring of the incentive – the cognitive framing effect introduced above (Fehr, Klein, and Schmidt, 2007: 141).

The ideas presented in this section seem to point to a particular understanding of the employment relation as based on a non-selfish or pro-social fairness motives driven by a particular understanding of fairness as reciprocity. Fairness in the employment relation, then, is contingent upon the actions of both parties, just as it is contingent upon the behavior of others, that is, the actions of a reciprocal employer may be less efficient if they are not met by employees with a reciprocal mindset, just as the employees' desire to reciprocate may decrease if they observe that those around them do not adhere to this norm set. Furthermore, this framework also draws attention the motives on which

the employment relation is based, just as it draws attention to the emotional foundation of the employment relation. Endogenously to the vertical relation, the pro-social behavior of the parties is contingent upon how the parties towards one another, and exogenous to the vertical relation, the pro-social behavior is based on the mindsets of other employees, just as it is also influenced by the fairness of their vertical interaction. In that sense, the framework described here allows one to push past the idea of employment relations as based on rational selfish motives of dispassionate social agents. To some extent then, this perspective infuses a sense of *life* into the theories on employment relations, just as it draws the attention towards its social foundation. Based on a similar foundation, Dan Ariely (2009) argues that the employment relation is based on two sets of relational norms, social and economic. Social norms are those on which community is centered, they are mutually beneficial and does not require instant payback. Economic norms, on the other hand, require, more or less, instant payback and equitable exchanges (Ariely, 2009: 68; Clark, 1984: 550). To exemplify this, consider Ariely and James Heyman's (2004) two games:

Figure 17: Illustration of Ariely and Heyman's (2004) games.



In both games, the participants in an experiment were asked to draw as many circles on a computer as they could within a five minutes period. In the game on the left hand side (A), one group of participants was offered 5 dollars, another 50 cents, and a third group was just asked to do it as a favor. The two groups who were offered money, supposedly, based their actions on an economic norm set and drew on average 159 and 101 circles, respectively. The third group, who was offered nothing at all, drew 168 circles on average (Ariely and Heyman, 2004: 788; Ariely, 2009: 71). In that sense, invoking actions

based on social norms seems to be more efficient than actions based on economic norms – this, in other words, seems to support the perspective, introduced above, that pure trust contracts invoke voluntary cooperation (pro-social behavior), whereas the introduction of incentives invokes an economic norm set which might undermine this voluntary cooperation. The game on the right hand side (B), is a pure social norms game, the game aimed at studying whether expensive gifts produced higher effort than inexpensive gifts, as it was the case with the difference in effort level between the 5 dollars and 50 cents offers (Ariely and Heyman, 2004: 790). The group of participants offered the inexpensive chocolate drew on average 162 circles, the group offered the expensive chocolate 169, and the third group asked to do it as a favor drew on average 168 circles, hence Ariely (2009: 72-73) suggests that gifts do not invoke economic norms even when given as payments for services rendered. In a third study, the participants were offered a “50-cent chocolate” and a “5-dollar chocolate” instead and it turned out that the effort levels chosen by the participants when the gift was explicitly priced invoked economic norms instead of social norms (Ariely and Heyman, 2004: 791; Ariely, 2009: 73).

Just mentioning the monetary value of a gift, then, seems to have a priming effect which invokes a different set of relational norms and behavioral patterns (Kahneman, 2011: 55). Kathleen Vohs, Nicole Mead, and Miranda Goode’s (2006: 1155, 1156) original study of the priming effect of money demonstrates that money-primed people are more self-sufficient, in the sense that they prefer to work alone and are drawn to tasks relying on individual rather than team performance; they are less likely to ask for help, less likely to help others, prefers isolation (they prefer others to keep a physical distance) – in short, money-primed individuals are more like the one’s described in neo-classical economics – or the agents described by Hobbes in the previous chapter. This also seems to support the perspective introduced above that the introduction of incentives changes the underlying structure of the game, because fairness concerns are substituted by selfish concerns. This, however, is also interesting in regard to the horizontal relation employee-colleagues, because approaches to reward management which emphasizes not only the individual performance, but also money may, as a consequence, undermine the social norms on which horizontal relations are based. Additionally, the introduction of performance-related-pay and, in general, an emphasis on the monetary

components may have adverse effects on the relation between the organization and the employee, because it invokes economic concerns and behavior governed by economic norms, hence as Ariely (2009: 82) argues, organizations wanting to generate employee relations based on social norms should rely more on non-economic benefits than economic incentives.

In contrast to the theories on fairness discussed in the previous chapter, the understanding of fairness as pro-social behavior entails a social conceptualization of fairness as how the parties in a relation acts towards one another (social interaction), just as it also emphasizes the fact that fairness is socially constructed in the relation. Although the relational approach focusing on the socioemotional dimension of inclusion/exclusion (see section 3.2 in chapter III) seems closely related with the perspective developed here, it should be noticed that the conception of fairness as pro-social behavior is closely related what I called social rationality in the previous chapter and, as such, it may be, but need not be, based on the Humean conception of sympathy as related with the emotional experience of inclusion/exclusion associated with esteem/self-esteem, see the discussion of Hume's concept of sympathy in chapter III. In other words, the question is not so much whether or not one feels included or excluded, rather the question is whether or not the actions of the other are equivalent in kind (fairness as reciprocity) or, at least, in accordance with the relational social norm which governs the particular interaction. Reciprocity as a basic principle of fairness (meta-theoretically defined as pro-social behavior or behavior in accordance with social norms) is contingent upon the behavior of the other party – in a sense then it is influenced externally by how the other agent acts toward oneself. From a Kant-Plato perspective, this would not matter, insofar as the agent should base her actions upon her principles and these principles should be based on reasons produced by an autonomous will and, as such, one is not excused if one replies unkindly on a unkind gesture, because this would, as recognized by Hobbes and Axelrod, lead to vicious circle which eventually would lead to the termination of the relation. Of course, whereas the Hobbesian agent is assumed rational, calculative and strategic, the human described here is not strategic, but reciprocates even when it is not in her interest to do so. Hence, whereas the Hobbesian agent is driven by instrumental concerns, the agent described in this chapter is driven by non-instrumental concerns related with the social normative foundation. From this it also

seems to follow that the perspective developed here takes into consideration the motivational dimension, insofar as it focus on why social agents espouses certain courses of actions instead of treating fairness solely as the outcome of an exchange, just as it seems to push the conceptualization of fairness close to the concept of social rationality. The problem might, however, be that the conception of rationality which is invoked here is, as it was argued in the previous chapter, too narrow in scope insofar as it is more to resemble with meta-theoretical critique of the assumptions on human nature inherent to neo-classical economics than an actual critique of rationality. Again, it might be argued that this (over)simplification of the conceptualization of rationality produces a set of theories which loose, at least to some extent, their intellectual roots in the actual relation between rationality and morality. Nevertheless, the perspective adds some important insight to, for example, the Kantian notion of rationality, insofar as external influence may not only negatively influence the agent's will (in the sense that it does not spring autonomous from a pure will – a pure intrinsic desire) it might also negatively influence social relations. That is, extrinsic influence may not only influence the agent's will, but also her sense of moral/social obligation because it alters the underlying nature of the social relation.

### **4.3 On Authority and Renegotiation**

In accordance with the discussions in the previous chapter, what has been argued so far in this chapter is part of a similar line of reasoning, although the distinction between instrumental and social rationality is substituted by a distinction between economic rationality and reciprocity or between the activation of economic or social norms. As such, the matter of fairness also becomes a matter of understanding motives and emotions involved in the exchange, which, at least to some extent, generates the underlying social dynamics of the employment relation. This perspective, however, is not only opposed to the perspective described in the neoclassical approaches to the employment relation, but also Simon's (1951) and David Marsden's (1999) theory inspired by Ronald Coase's (1937) and Simon's (1951) respective theories. Although Simon (1991), later reformulated his theory on the employment relation to encompass ideas about loyalty and commitment, as argued above, his original theory on the employment

relation was inspired by Barnard's (1938/1968: 168) idea about the *area of acceptance* describing the area of the employee's opportunity set over which the organization has authority – that is, by engaging the employment contract, the employee accepts, within some limits, the organization's right to hold authority over some of her actions (Simon, 1951: 294, 295; 1945/1997: 10). The idea about the relation between the employment contract and authority was originally introduced by Coase (1937: 388) who saw the authority relation as a natural consequence of the employment contract, and as an important way in which this differentiated itself from the relation between two independent contractors. According to Simon (1951: 296), this area of acceptance emerges from negotiations between the employee and employer and, as such, it represents an equilibrium state which *satisfices* both parties. On the one hand, this area of acceptance can be viewed as a logical consequence following from the non-specific nature of employment due to uncertainty with regard to future needs (Simon, 1991: 31), and, on the other hand, this area of acceptance also describes the rule based nature of the employment relation, that is, the employment relation can be viewed as negotiations of the rules on which this should be centered – a kind of dynamic equilibrium within, however, certain institutionalized limits (Marsden, 1999: 17, 21; 2000: 324, 326; 2004b: 663). The assignment of tasks, for example, can be thought of as a negotiated set of rules specifying which tasks *can* be assigned to a particular job within the organization.

The perspective on the reciprocal foundation of the employment relations, then, covers primarily what is often referred to as the psychological dimension of the employment contract, that is, the social normative foundation which lend credit to the *promissory nature* (the employee's promises (future intent) about, within her area of acceptance, to adhere to her obligations, and the employer's promise (warrant) on paying the employee for her services) on which the contract is based (Marsden, 2004b: 665; Rousseau, 1995: 9, 11, 13, 15). This dimension of the contractual agreement, then, is close to the one presented above, hence one can add to this that the social normative foundation of the contract is important, because: (1) it provides support to the promises implied in the economic dimension of the contract, just as it covers, by imposing obligations, some of the problems which might emerge due to information asymmetry and opportunism; and (2) not all aspects of social interaction can be covered by a formal contract (Marsden, 2004b: 670; Fehr, Fischbacher, and Gächter, 2001: 2). Indeed, as

also emphasized by Denise Rousseau (1995: 76), reciprocity is often assumed, thus deviations from this social norm may invoke strong emotions within the aggrieved agent who might be willing to, as argued by Rabin (1993), incur high costs when reciprocating the unkind act. The psychological dimension of the contract may also be supported by the economic dimension, because this might specify the kind of performance valued by the organization (Marsden, 2004b: 666, 670). The last dimension of the contract emphasized by Marsden (2004b: 668) is the legal dimension which provides the framework within which certain key aspects of the relation can be enforced by an impartial third party. In that sense, to understand the institutional framework within which this area of acceptance is negotiated, one needs to take into account these three dimensions of the contract.

Furthermore, it might also be argued that the employment relation is centered upon two additional conditions: (1) the employer will not try to gain authority over actions outside the area of acceptance, and neither will she take on actions which deliberately decreases the utility of the employee; and (2) the employee will accept the employer's authority within the area of acceptance, and choose the level effort needed to fulfill the contract. Viewing the employment relation as one based on authority should, however, not be confused with an argument for subordination, because as Marsden (2004a: 352; 2010: 202) argues, it draws attention away from the underlying nature of the contract as a negotiated agreement, implying that one should not neglect the employee's voice in renegotiations of the relational terms – albeit as he argues in a recent article, individual voice may be a risky business due to relational asymmetry implying that in many cases the employee must rely on exit, which is only credible if the employee intends to leave and she is able to obtain a position at another organization, or loyalty (Marsden, 2013: 224). In renegotiations of this area of acceptance, one should not, then, neglect the employees' bargaining power, just as one should not neglect the importance of agreement – that is, without the employees' acceptance, sustainable vertical relational ties in the relations employee-management and employee-organization are problematic (Marsden, 2010: 202). This perspective enables Marsden (2004: 354, 359; 2010: 188) to explain the seemingly paradoxical nature of the introduction of performance-related pay in the British public sector, because performance improvements was obtained by other means

than increased motivation – namely, because of goal-setting processes which facilitated goal congruence between the individuals' and the organizations' goals.

#### **4.4 Authority, Incentives, and Non-economic Motivation**

Authority, in form of pressure of the employees might, however, have adverse effects because it might undermine the relation, insofar as it undermines the employees' motivation and their confidence in the fairness of management or organizational procedures, just as it might undermine the intrinsic motivational dimension of the job which in some cases may keep the employees from shirking, in the sense that if the employees take pleasure or pride in their job they will be less prone to perform actions which in some form may undermine this foundational sentimental structure (Marsden, 2004a: 361). In regard to this, Marsden (2010) comments on the introduction of performance-related pay in the British public sector:

“...if management pressure employees too much, then some of the safeguards against abuse that arise from the intrinsic value of the work, such as beliefs in providing a public service, can be undermined. In the Employment Service, it could be argued that these beliefs held potentially opportunistic behavior in check – if you believe your job is to help people then there is no point in faking the numbers. If the intrinsic value is downgraded, and the employees are also penalized for giving attention to that rather than sustaining the desired case throughput, then the system can tip out of control...” (Marsden, 2010: 198)

As Marsden (2010) seems to argue in this quotation, the employees' potential opportunistic behavior may also be reduced due to public service motivation or pride of workmanship. On the latter, W. Edwards Deming (1982/2000) argues:

“Piece work is more devastating than work standards. Incentive pay is piece work. The hourly worker on piece work soon learns that she gets paid for making defective items and scrap – the more defectives she turns out, the higher her pay for the day. Where is her pride of workmanship?” (Deming, 1982/2000: 72)

In that sense, besides social motives or social preferences described above, one may also add intrinsic motivation as a factor which influences the employment relation. In contrast to the perspective discussed in the first part of this chapter, the altruistic motive implied in public service motivation is not inherent to the dyadic relation; rather it is

directed towards some third party – in public management, the client (Andersen, 2012: 3). Public service motivation may be defined as follows:

“...an individual’s orientation to delivering service to people with the purpose of doing good for others and society.” (Hondegem and Perry, 2009: 6)

The motive, then, seems to be altruistic in nature, although this affective dimension of public service motivation was regarded by James Perry and Lois R. Wise (1990:369) in their seminal article, as the least important.<sup>30</sup> Lotte Bøgh Andersen and Thomas Pallesen (2008: 31) demonstrates in a study how public service motivation may be viewed as a supplement to theories on intrinsic motivation, because these do not take into account, in a similar way as public service motivation, the intrinsic motivation one may enjoy from, for example, helping others; rather these theories focus on task-based motivation (Andersen, 2012: 7). As argued by, for example, Bruno Frey (1997: 18), economic incentives may lead to a crowding-out effect of intrinsic work motivation if the incentive is perceived as controlling. Andersen and Pallesen (2008: 41, 42) find a positive relation between the number of publications at Danish universities and economic incentives, however, it also matters whether this is perceived as supportive or controlling – hence, the simplified results are that when economic incentives are perceived as supportive the effort increases, whereas in the opposite situation in might lead to a decrease in effort which might be viewed as a crowding-out of public service motivation. In other words, the intention matters. Furthermore, as argued by Marsden (2004a; 2010), this out-crowding of motivation may not only undermine the relation between the organization and its employees, but also its fair procedures.

#### **4.5 Tentative Conclusions**

The aim of this chapter has been to study how the functionality of a compensation system based on individuality – pay-for-performance – is influenced by the interaction between the horizontal and vertical dimension if fairness is defined as reciprocity. In that sense, the perspective discussed here continues the primary distinction discussed in the

---

<sup>30</sup> The authors also mention the rational-based motives, for example, participation in policy processes or commitment due to identification with the program’s aim; and the norm-based motives, for example, desire to serve public interest or duty towards society (Perry and Wise, 1990: 368)

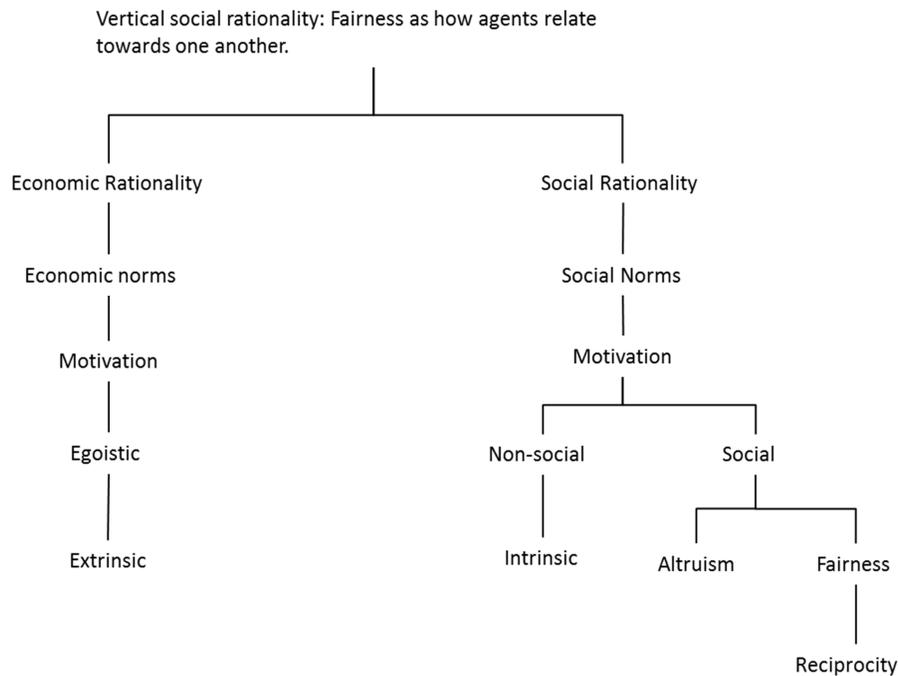
previous chapter between rationality and social rationality. In the previous chapter, the focus was primarily on the relation between rationality and fairness with regard to the horizontal dimension, whereas in this chapter the vertical dimension has been emphasized. Reciprocity may be defined as based on the principle of fairness that actions are responded to in an equivalent manner – that is, kindness is responded to in a kind manner, and unkindness in an unkind manner. However, reciprocity cannot be reduced to a matter of actions already performed, insofar as intentions matter – that is, the intention behind the action matters, at least in comparison with its alternative. Moreover, in this chapter, fairness was associated with how agents act toward one another – a definition which is closely related with the definition of social rationality discussed in the previous chapter.

Reciprocity may, then, be viewed as a kind of social rationality, insofar as it is related with the social dimension of interaction occurring within the relations, employee-management, and employee-organization. Furthermore, because reciprocity is accompanied by the motivation to reciprocate in a like manner, it may restructure the underlying preferences of incentives games insofar as the incentive is associated with a particular intention. If the population of reciprocal minded individuals is sufficiently large, the organization may be successful in producing positive effects when offering a compensation level above the market rate, insofar as this may be treated as a gift. If, on the other hand, the incentive is framed as a matter of control or not perceived as equivalent to the actions performed by the employee, it might undermine intrinsic motivation, public service motivation, and the employment relation itself.

In addition, some studies show that the very introduction of monetary incentives might undermine the social relational nature of the employment relation, insofar as it changes the governing norms from social to economic norms – hence, a social out-crowding effect can be identified in the sense that pro-social behavior may be out-crowded. Hence, the positive benefits of reciprocity might be undermined, just as the horizontal relations might be influenced, insofar as the activation of an economic norm set activates behavior associated with economic norms and, as such, it might undermine the social relational nature of the horizontal dimension. Hence, it might reduce social rationality to a matter of cost and benefits.

However, just like one analytically should balance rationality and social rationality as the foundation of fairness, one should also balance social rationality with the obligations inherent to the employment relation. That is, while reciprocity covers the social dynamics on which the employment relation is based, one should not neglect that the employment relation is a particular kind of social relation. Thus, the structure of the vertical relations cannot be reduced entirely to matters of social rationality, just like horizontal relations cannot be reduced to a matter of rationality. Nevertheless, this should not be interpreted as a defense of authority; rather it should be viewed as a matter of capturing the distinct nature and complexity of vertical relations within an organizational context. Based on the discussions, the following figure can be produced:

Figure 18



The left hand side of the figure, is similar to the Hobbes-Hume-Bentham tradition to rationality identified in the previous chapter. Neither of the classical authors, however, identified reciprocity as the foundation of morality, insofar as this would, probably, have been against the rational foundation, because it reduces social interaction to a matter of conditional behavior – interaction without reflection, principles, reason, or calculation. From a non-instrumental perspective, vertical relations could be explained based on a Kantian perspective, perhaps combined with the perspective proposed by Deci and

Ryan (2000). In that sense, reciprocity may be viewed as a particular kind of social rationality, insofar as it must be assumed that the agent who applies reciprocity as the basic principle of fairness in social interaction also believes that it applies, just as it seems to reduce social interaction to a non-reflective matter and, as such, it might be considered as more social and less rational than the perspectives on social rationality discussed in the previous chapter. To reduce interaction to non-reflective behavior based on emotions which follows from the attribution of intentions may, of course, be viewed as a reductionist account of vertical relations, insofar as it might understate the implications of formal contractual relations, because the employee cannot just reciprocate perceived unkindness, at least not without it having consequences that employee must be willing to face. In a similar vein, reciprocity might also lead to vicious circles and seems, as such, to neglect the Plato-Kant notion of fairness in which the agent is obligated to govern herself, hence she cannot excuse her own unkind behavior with a reference to the unkindness of others – the very notion of agency implies choice and the possibility of authenticity – a will and a notion of the self which is not contingent upon extrinsic factors. The model proposed in this chapter, nevertheless, draws attention to an important feature of non-consequentialistic behavior, because it emphasizes that behavior is not contingent upon outcomes, but on how one is treated – and, as such, the social dimension of vertical relations becomes explicated in the fact that behavior is socially conditioned in the sense that it is a reaction on other social agents' behavior and not some exogenous factor such as money. That is, it takes seriously that, for example, compensation occurs within a social interaction – which unfolds itself within a social relation – and, as such, it can be treated, analytically, as an action which like other social gestures carries with it a particular symbolic notion – it is impregnated with meaning – just as it must assumed to be rationally selected to signify something in particular – the intentional dimension. In a similar manner, it might also be argued that actions equivalent in kind must also be considered as the result of some process of choice, thus some kind of decision process must be assumed to precede the actual choice. Framed in this perspective, the cognitive element cannot be reduced to a matter of mere reciprocity, thus to frame reciprocity as irrationality would be to neglect this process of choice and the cognitive link.

## Chapter V

### Fairness and the Relational Status Quo

#### 5.1 Introduction

In the previous chapter, I discussed the normative foundation of vertical relations and identified two sources to social rationality, fairness considerations in interaction and formal obligations inherent to the employment contract. The vertical relation is a particular kind of relation, in the sense that it is a relation between asymmetric partners, just as it is governed by a formal contract as well as a social contract. In juxtaposition to this, the horizontal relation is solely centered upon an informal social contract and, as such, it is more vulnerable than the vertical relations because if the social normative foundation, when agents ceases to take into accounts how their actions might affect the lives of other social agents, the Hobbesian state of war might be realized. The opposite of social rationality may, in that sense, not be strategic reasoning, but indifference, because as long as one cares about one's own fairness one still needs to take into account the social dimension of realizing this. That is, as also argued by Folger (2012), in order to understand fairness as morally based one needs to take into account the self-constituted obligation, as argued by consequentialists the obligation which follows from one's actions having consequences, or as I argued in chapter III, as a synthesis, the obligation which follows *naturally* from taking part in social life, from being influenced and from being allowed to influence the lives of others.

Conceptualized in this manner, neither horizontal nor vertical relations can be understood without taking account of the normative foundation, thus while libertarian inspired theories, such as Lawler's (1990; 2003) theory of compensation discussed in chapter II, emphasizes freedom (unboundedness), tolerance of individuality and respect of difference, they seems to neglect the normative dimension of social life – the fact that social agents hold normative expectations on what signifies a fair treatment, just as they also tries to act fair towards other social agents. That is to say, libertarian inspired theories put forward a conceptualization of relations as centered upon a non-social understanding of justice as being related with how the organization treats free agents in a just

manner and, in that sense, the relation which emerges may be fair, but based on an unbounded conception of justice – a non-social understanding – in which the respect of individual freedom has generated a portrait of organizations as non-social entities centered solely upon necessary relations characterized solely by necessary obligations. Social obligations emerging from engaging social relations may, then, be viewed as a self-commitment not following directly from the relation and as irrelevant. Hence, the modern conceptualization of organizational relations described in chapter II may be viewed as producing personalized vertical relations in the sense that these are centered upon respect of individuality, that is, it produces non-social relations – personalized relations but not *personal* relations.

From this it seems to follow that the central question of distributive fairness cannot be whether equality or equity is the fairest principle; rather the question seems to be a matter of the fairness of the action and not its outcome – that is, if equity is perceived as the fairest principle it must be defensible towards oneself and towards others – from Plato-Kant perspective, one must want the principle to be applied in the determination of one's own outcome as well as the outcome of others. In some cases, as argued by Boudon (2001), a distribution is not accepted because of value congruence (commitment to the *right* values) or because it is technically fair based on an assessment of the system's ability to produce accurate measures, but solely because one is not able to produce strong reasons against it not being fair. Nevertheless, in accordance with the discussion of social rationality in chapter III, one might view this emerging from a balance between what is fair to oneself and what is fair to the other. The aim of this chapter is to discuss this balance and how it is relationally constituted and influenced by social norms and social emotions such as envy and guilt. The point of departure is equity theory, because these theories allows one to study how holding the distributive norm of equity influences the agent's perceived right to claim vertically as well as how it is horizontally justified.

Aristotle (350BC/2007: V.3, 1131a/127) distinguishes between distributive and corrective fairness. Distributive fairness concerns the distribution of goods between similar (or comparable) individuals, whereas corrective fairness concerns corrections or compensation which restores equality in daily matters brought about by the distribution

of goods. Although the distribution occurs between similar individuals it is not necessarily equally distributed, because Aristotle acknowledges that equal distributions between non-equals is just as problematic as unequal distributions between equals. A just distribution taking differences in *worth* (that which differentiates and offsets equality) into consideration is, then, proportional – and more specifically, it is *geometric proportional*<sup>31</sup>. Corrective fairness, on the other hand, is not based on geometric proportions, but on *arithmetic proportions*<sup>32</sup> to offset the proportional differences implied in a geometric proportional distribution – in that sense, corrective fairness does not take differences into consideration, like *Justitia*, the goddess of justice from Roman mythology, who is often, portrayed wearing a blindfold and holding a scale. Corrective fairness may be viewed as a means against *too little* (adding) or *too much* (subtracting), in the sense that corrective fairness has as its aim to restore balance in a state of imbalance (Aristotle, 350BC/2007: V.4, 1132a/130). Those who restore balance are the impartial judges (a third party) and not the parties themselves as it is the case with reciprocity which I discussed in the previous chapter.

Fairness, as discussed so far, is not measured in absolutes in the sense that, for example, dividing 6 units between two equal social agents is not fair, because each agent receives  $(6/2)$  3 units, rather it is fair because each receives *a half* – the half then stands in the same relative relation to the whole, as each agent to the other, as an equal relative part of the whole. The distribution is fair because it reflects the relative relation between the two social agents, thus each should receive  $\frac{1}{2}$  of each unit. If the two agents are not equal, then, the number of social agents is equal to the distributional weights, e.g. if A's worth is twice of B's, then  $A = (2:4)$  and  $B = (1:2)$  such that each distribution stands in a relation to the whole, as each agent stands in relation to each other – reformulated as a geometrical proportional relation;  $2:4::1:2$  becomes  $2/1 = 4/2$ . If A is not just similar, but also equal to B, then both should receive half of the whole, whereas if they are similar but unequal, then they should receive a part of the whole equal to the

---

<sup>31</sup> A geometrical proportional distribution can be illustrated as follows. Consider two squares: (A,B,C,D) and (a,b,c,d). The second square is similar to the first, just smaller or reduced. That is, A:a, B:b, C:c, and D:d. If the largest square is 2 times as big as the smaller one, then, 8 is to 4 as 16 is to 8 and so forth (Smith, 1923/1951:63).

<sup>32</sup> An arithmetic proportional distribution is different from the geometrical and can be illustrated as follows:  $B-A = D-C$ . So, the two squares mentioned above are now equal (and not just similar) meaning that  $4-8 = 4-8$  (Smith, 1923/1951:63).

relative difference between them – in other words, the relational asymmetry is held constant such that the inequality  $A > B$  is the same before and after the distribution occurred. The basic focus of this chapter can, then, be illustrated in the following manner:

Figure 19: Cause and Effect Diagram.



The figure above illustrates the causal relation which is the primary concern of this chapter. Relational concerns are, here, viewed to be the cause of fairness, albeit these are mediated by another mechanism, for example, social preferences – illustrated as circle because it is process and not just as mediating variable (see my previous discussion of social mechanisms in chapter I). Aristotle argued that relational status quo concerns have an impact on whether a distribution is found to be fair or unfair. That is, the initial situation matters, in the sense that this initial situation is a *psychological reference point*. This entails that if the parties are, in the initial situation, unequal, then, an unequal distribution which maintains the initial proportional difference is found to be fair, whereas equal parties found an equal distribution to be fair. Aristotle (350BC/2007) focused, then, only on the social preference, maintaining relational status quo, because he considered all deviations from this as unfair. That is, a social agent who for some reason would prefer a distribution not in accordance with this proportional status quo would either be unfair to herself (because the agent receives too many burdens or too little goods); or unfair to the other (because the agent receives too few burdens or too many goods). Fairness, then, lies in the middle between what is unfair to oneself and what is unfair to the other party – see the illustration below (Aristotle, 350BC/2007: V.5, 1133b/134):

Figure 20: Illustration of Aristotle's idea of fairness as the middle.



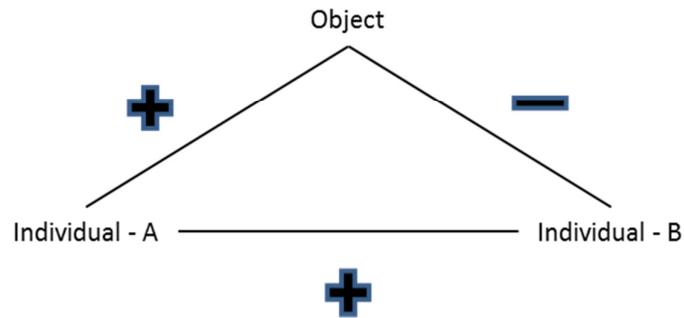
As the continuum illustrates, the target of fairness is not just the individual herself but also the other. The fair action springs from a rational foundation in the sense that the foundation is practical rationality and concerns reason as well as passion (affection)

(Aristotle, 350BC/2007: VI.2, 1139a/150). The central question of fairness, thus become; “is that which is fair to you also fair to me?” and equivalently “is that which is fair to me also fair to you?”. These two possible phrases of fairness explicate the underlying logic of fairness which is: (1) *relational* in the sense that (1.a) it is centered between two extremes (“unfair to me” and “unfair to you”); and (1.b) as a relational characteristic of the structured relation between two social agents (1.b.a); and (1.b.a) their interaction; (2) *comparability* in the sense that (2.a) it exists between two *similar* parties; and (2.b) it has character of an *average* or a *compromise* between the two extremes (“that which is fair to me, but not to you” and “that which is fair to you, but not to me”); and (3) the aim of fairness is not the happiness of either party in isolation, but rather the community, insofar as fairness secures harmonious co-existence, this, however, is not contrary to individual happiness, in the sense that the fair action is in accordance with practical reason (reason and morality) – that is, in accordance with the principles of the good life which is held by all rational individuals (Aristotle, 350BC/2007: I.7, 1097a-b/38-39; V.1, 1129b/124).

## **5.2 The Balance, the Middle, and Cognitive Dissonance**

The metaphor of the fair as the middle becomes in the early modern models on equity theory illustrated based on a balance metaphor inspired by Fritz Heider’s (1946) cognitive balance theory. Individuals, he argues, cannot simultaneously hold two contradictory attitudes towards another individual – one cannot simultaneously love and hate another individual, just as *A* might have a positive relation to *B* but not to *C*, and if *B* also has a negative relation to *C*, then: (A’s negative relation to C) + (B’s negative relation to C) + (A and B’s positive relation to one another) = relational balance (Heider, 1946: 111). What Heider (1946) argues is, in other words, that attitudes are transitive relations and that relations characterized by emotions like envy and competition has implications on this – *A*, for example, cannot simultaneously desire to obtain some object and keep a positive relation with *B*, but not desiring *B* to obtain the object.

Figure 21: The implications of emotions on relational transitivity.

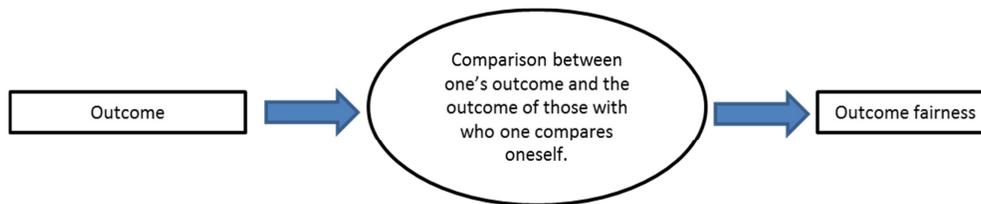


This also entails, as Homans (1958) argues, that *A* cannot believe her own part of a distribution to be unfair, while believing that *B*'s part of the distribution is fair, and the relational proportionality to have been maintained. To Homans (1958: 603-604) this implies rank equality or *status congruence* – that is, if *A* and *B* have different *costs* (input) then, they should also be paid different rewards to maintain an equal profit (Profit = reward – cost). That is to say, *A*'s profit should stand in relation to her input, as *B*'s profit stand in relation to his input – a reformulation of the Aristotelian principle of proportional equality. Proportional equality, then, states that: equal costs should be reflected as equal rewards, while unequal costs should be reflected in the rewards in a manner in which maintains the proportional difference in costs is reflected as a similar proportional difference in rewards so as to maintain the relational proportionality.

Besides this mechanism in which the relational status quo is used as a reference point for fairness judgments, this theory also put forth another aspect of fairness which was also, albeit rather vaguely, implied by Aristotle, fairness is a judgment based on social comparison and as such it is relative to the context in which the comparison takes place. How fair individual *A* finds her profit to be is conditional on three aspects: 1) how fair she finds *B*'s profit; 2) her perception of the proportional relation between herself and *B* in terms of costs (or input); and 3) how her profit compares with that of *B* (Festinger, 1954). Now, as Aristotle also implied, proportionality requires similarity (or likeness) but not equality, hence an average performer may choose to compare her profit with that of a similar other, and not that of high performer (Festinger, 1954: 120). Social comparison is used as a *proxy* in situations in which no objective measure exist (Festinger, 1954: 118), and as such if *A* finds *B*'s profit to be fair, but the proportionality of her profit compared to that of *B* to be unaligned with her perception on the propor-

tionality of their respective input, A will experience *relative deprivation* – a situation in which she may believe herself to be equally or more entitled to some reward than B (Merton, 1938/1968: 186). Implying that the deprivation one experiences from something is relative to the deprivation one experiences from those with who one compares oneself, fairness becomes, here, a matter of the social norms on which the situation is centered – in the sense that social norms on what to desire/ the goals of one’s actions (values) is often associated with social norms on how to legitimately obtain these socially legitimate objects (instrumental norms) (Merton, 1938/1968: 190). The model below illustrates the effect.

Figure 22: Social Comparison as a Psychological Mechanism.



The impact of social norms on pay systems was not only recognized by social psychologists and sociologists, the idea was also adopted by two of the founders of modern theories on pay systems, John Hicks (1974) and Elliott Jaques (1961). According to Hicks (1974), the normative foundation of pay systems is problematic, albeit a necessity. This is also illustrated in the following two quotations:

“...system of wages should be established so that it has the sanction of custom. It then becomes what is expected; and (admittedly on a low level of fairness) what is expected is fair.” (Hicks, 1974: 65)

He continues this argument later in his work:

“...it is bound to work more easily if it is allowed to acquire, to some degree, the sanction of custom – if it is not, at frequent intervals, being torn up by its roots.” (Hicks, 1974: 79)

A pay system should, according to this perspective, be based on social norms - perceptions of fairness – because like any other market system it is shaped by expectations. Hicks (1974) is here, probably, arguing against John Maynard Keynes’ (1936/2008: chapter 2) theory on the psychology of the expectations and how employees care about what other employees in another organization are paid. Jaques (1961), on the other

hand, does not share this concern and like most equity theorists he considers the drive of equitability in the employment relation to be based on fairness concerns. Expectations plays here, also, a pivotal role, because these are produced by the state of the general economic situation, in the sense that a positive economic situation produces high expectations which increases the demand for equitable relations – Jaques (1961:148) is, here, not referring to this in a *real* economic sense. Equity describes the *felt-fairness* of differentiated payments and, as such, it concerns the relative treatment of one employee compared to another (Jaques, 1961: 148). Equity, Jaques (1961) argues, allows the “...*the worth of a man’s work to be recognized.*” (Jaques, 1961: 175). He argues further on this point:

“Because of differences in capacity, not everyone can make an equal contribution to the final result. Not everyone should share equally in the rewards. There is a conception of each man reaping the just rewards of his own labour which is deep-seated in the mind.” (Jaques, 1961: 175)

Jaques (1961), unlike Homans (1958) and Adams (1963), whose perspective I will present in a short while, does not perceive equity as a matter of maintaining some proportional relation between comparable others, but focus primarily on equitable employment relations – the relation between time-span<sup>33</sup> and money (Jaques, 1961: 145). Equitable payment, although not dependent upon it, can be viewed as relative to the employee’s actual payment. If the equitable payment is equal to the actual payment, the employee will perceive her pay as reasonable. If the equitable payment is higher than the actual payment, the employee is dissatisfied, and if the actual payment is higher than the equitable payment the employee will feel that he “...*is paid within a higher range than he can ever hope to maintain.*” (Jaques, 1961: 154). The intensity of the experience varies with the discrepancy between the equitable and the actual payment. Moreover, a  $\pm 3$  percent deviation is considered to be within the reasonable frame, while a 5 percent below equity deviation is considered unfair, at a 10 percent below equity deviation the employee considers leaving her job, and at 15-20 percent below equity deviations the employee is looking for a new job (Jaques, 1961: 154). Whether or not the employee actually leaves the organization is, however, dependent upon the economic situation, because in recession periods, the employee is more likely to put up with her situation,

---

<sup>33</sup> Time-span is a term developed by Jaques (1961:52) which measures the degree of responsibility held by the employee.

with psychological costs such as depression or lack of motivation to follow (Jaques, 1961: 155).

A similar logic, as just described in relation to under-compensation, applies to situations of felt over-compensation. At a 5 percent over-compensation, the employee feels that she is getting more than her fair share, and at a 10 percent over-compensation, the employee feels anxiety (Jaques, 1961: 155). Jaques (1961) adds to this:

“He experiences feelings of guilt with regard to others who are not doing so well. But guilt may be warded off by a devil-take-hindmost attitude. He may express resistance to change in the content of his work, to the introduction of new methods, or to transfer to new jobs. Greed and avarice may be stimulated, with a resulting anti-social grasping for further relative gain regardless of the consequences for the common good.” (Jaques, 1961: 155)

Jaques (1961) seems, here, to be arguing that over-compensation, like under-compensation, can lead to what Merton (1938/1968) called, borrowing Emile Durkheim’s (1897/1951: 242) term, a state of anomie (normlessness) – a state in which the individual retracts herself from society, in the sense that she stops conforming to the norms of society. Over-compensation, then, leads to a state of psychological anxiety caused by a feeling of guilt combined with a sense of undeserving or unworthiness – and if the individual stays in this state too long her personality and moral will be crowded-out by greed and egoism. Under-compensation, on the other hand, may have a similar effect brought on by a sense of relative deprivation, which might lead to a state of hopelessness and unworthiness.

Adams (1963: 426) following Jaques’ (1961) reasoning, argues that both over-compensation and under-compensation may call forth within the individual a state of psychological anxiety, although the anxiety of under-compensation is worse, in the sense that the threshold for unfavorable inequity is lower than for favorable inequity which the individual might rationalize to herself. In contrast to Jaques (1961), Adams (1963: 426) argues that social agents have different thresholds of inequity which imply that the emotional intensity with which the sense of inequity is invoked within the individual differs from person to person – this point is also emphasized and further developed by equity sensitivity theorists.

Adams' (1963: 423) theory on inequity is inspired by Festinger's (1957/1985: 3-4) theory on cognitive dissonance stating that individuals who experience a difference between how "the world should be" and how "the world actually is" will try to reduce this state of cognitive dissonance – like Festinger's (1954) theory on social comparison and Heider's (1946) theory on cognitive organization, this theory is based on a balance assumption, the state of *consonance*. According to Adams (1963: 423-424), inequity relates to a felt imbalance between input (e.g. education, experience, training, skill, seniority, effort etc.) and outcome (e.g. pay, intrinsic rewards, different kinds of benefits, job and social status etc.), these input and outcomes must be recognized by the agent conducting the fairness judgment as relevant to the exchange or else they will represent no psychological value to the individual, compared to that of another individual. The intensity of the sense of inequity, then, is relative to the magnitude of the discrepancy expressed as a proportional difference, thus the intensity of the sense of inequity is worse in the first case than in the second case showed below (Adams, 1963: 426):

$$\text{First case: } \frac{\text{Outcome}_{\text{Person}}(\text{Low})}{\text{Input}_{\text{Person}}(\text{High})} < \frac{\text{Outcome}_{\text{Person}}(\text{High})}{\text{Input}_{\text{Person}}(\text{Low})}$$

$$\text{Second case: } \frac{\text{Outcome}_{\text{Person}}(\text{Low})}{\text{Input}_{\text{Person}}(\text{High})} < \frac{\text{Outcome}_{\text{Person}}(\text{High})}{\text{Input}_{\text{Person}}(\text{High})}$$

Because individuals are motivated to generate balance – and fairness is balance – they will take different actions to bring this state about (Adams, 1963: 427). In the first case, illustrated above, Person might try to reduce her input, because comparing the inputs, isolated, it can be seen that Person supplies, or believes she does, a higher level than Other does. This should be viewed in relation to Person's own outcome/input equation, hence the comparison takes place in the following manner – Person feels a sense of inequity in comparison with Other and focus, then, first on her own outcome/input equation, and secondly how the component which seems to bring about the sense of inequity, here input, compares with that of Other (Adams, 1963: 428):

$$\text{Identifying the component which causes imbalance: } \frac{\text{Outcome}_{\text{Person}}(\text{Low})}{\text{Input}_{\text{Person}}(\text{High})}$$

$$\text{Compare that component with Other's : } \text{Input}_{\text{Person}} > \text{Input}_{\text{Other}}$$

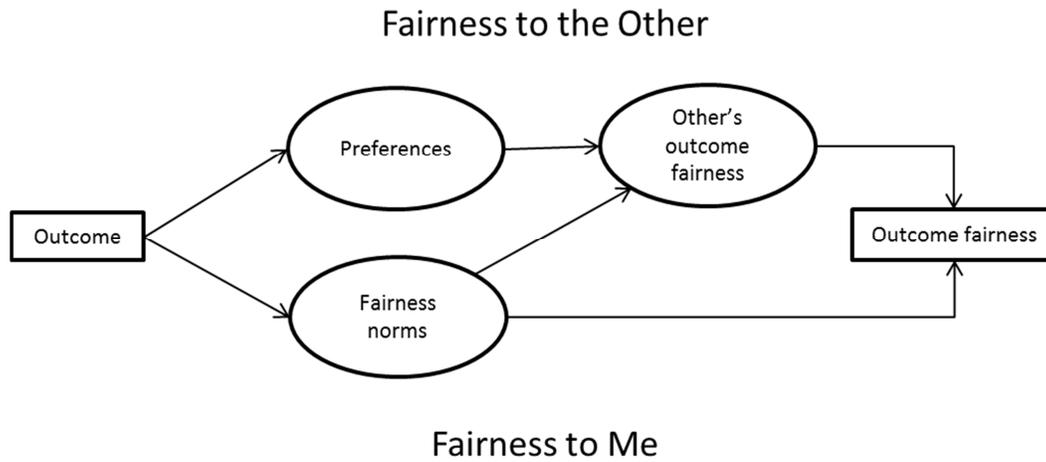
In the opposite situation, in a state of over-compensation, the individual will try to reduce the dissonance by, for example, donate to charity (Adams, 1963: 429). As emphasized by Aristotle (350BC/2007), Jaques (1961), and Adams (1963), people prefer equitable, fair, interaction – a claim which not only concerns their own gains, but also the gains of other people. Messick and Sentis (1979: 419) problematize this assumption by asking if what is found fair is also preferred, and on which do people act, values or preferences? To study this, the authors study three different situations, one in which the outcome of Other has been assigned and the respondent is to assign an outcome to herself, a second in which the respondent's outcome has been assigned and she is to assign an outcome to Other, and a third situation in which the respondent is to split some amount between herself and Other (Messick and Sentis, 1979: 422). Furthermore to study the impartiality – independent of self-interest – of fairness judgments, implied by the authors above, Messick and Sentis (1979: 428) designed symmetric cases – hence in the first case Other's input was higher than the respondent's, and in the second case the respondent's input was higher than Other's. In the first situation, when they were to assign an outcome to themselves and their input was lower than Other's, the respondents found the equality rule to be most fair, while in the other case, when they were to assign an outcome to Other and their input was higher than Other's, they found equity rule to be most fair, and in the even split situation, the respondents found an even split most fair (Messick and Sentis, 1979: 429, 432). In other words, the respondents' fairness judgment, when they were to assign an outcome to themselves, was egocentrically biased.

Messick and Sentis' (1979: 430) study also shows that a majority of the respondents preferred a distribution which was different from the one they found fair, when they were to assign an outcome to themselves they distributed a higher amount. In the second case, almost all the respondents also preferred the distribution they judged to be fair, when they were to assign an outcome to Other. The same consistency between judgment and preference was also showed in the even-split case, albeit with less agreement. Furthermore, when asked which distribution they would choose, the majority of the respondents chose one similar to the one they judged to be fair (Messick and Sentis, 1979: 431). In the first case, when the respondents' were to assign an income to themselves, the majority of the respondents' choice was equal to the outcome they found

fair. In the second case, when the respondents' were to assign an outcome to Other, the following relation between judgment, preference, and choice applied. When the fairness judgment is equal to the preferred distribution, then, 93 percent of the times, the choice is the same as the fair distribution. When the fairness judgment is not equal to the preferred distribution, 58 percent of the times the respondent chose an outcome equal to the fair one, 28 percent of the time chose one similar to the preferred, and the rest of the time they chose some intermediate outcome between preference and the one they found fairest (Messick and Sentis, 1979: 433). In that sense, when assigning an outcome to themselves, the respondents displayed an egocentric biased tendency which from the fact that they chose the outcome they perceived as fair, while, in the second case, when assigning an outcome to Other, the respondents' choice reflected a mixture between the outcome they found fair and their preference. This implies, then, when social agents are to assign an outcome to other social agents their choice will reflect some mixture between what is found to be fair and preferences – this mixture is what has recently been referred to as social preferences.

What Messick and Sentis' (1979) perspective on fairness seems to be advocating is that when evaluating an outcome distribution, this evaluation process is a mixture of several mechanisms related to what I, here, call the two roads to fairness – a perspective inspired by Aristotle's distinction between "unfair to me" and "unfair to you". The social agent who is to evaluate the fairness of her private outcome will do so based on the self-serving fairness norm – e.g. equity or equality – invoked in the situation. When evaluating the fairness of Other's outcome, on the other hand, the agent's evaluation is a mixture of preferences and fairness norms. This is illustrated in the figure below:

Figure 23: The Two Roads to Outcome Fairness.



The procedure of justifying one's evaluation cannot, then, be based solely fairness norms or at least not in an *objective* sense. What Messick and Sentis (1979) also draw attention to is a distinction between, moral judgments, personal preferences, and the actualizations of these in choices. Firstly, they argue that the moral judgment is not fair in the sense that it is not impartial, but self-serving. And secondly, they argue that the respondents' preferences are not equal to their moral judgments nor their choices – implying a difference between “what I should” in a normative sense and “what I want” in a matter of what is most preferred. This distinction is neither present in Jaques' (1961) nor Adams' (1963) equity theory, but it is a core distinction in moral philosophy, and practical reasoning as argued in chapter III. Furthermore, what seems interesting here is also the fact that the authors find that the respondents' choice, primarily, reflects their fairness norms and that this is egocentrically biased. Based on Merton's idea about social norms as legitimizing goals and means, introduced above, one may argue that it is not socially legitimate to base one's choice of outcome distribution on one's preferences; hence fairness principles are invoked not, however, in a not *hypocritical* way, but rather as a legitimate way of reasoning about other possible distributions.

The aim of introducing these classic theories is, on the one hand, to enable an understanding of the modern theories on social preferences and, on the other hand, the aim is also to enable a reflexive understanding of the constitution of claim rights in vertical relations. In that sense, these classic theories on equity captures some basic features of

the cognitive constitution of these claims, insofar as these theories demonstrates that if equity is held as a fairness norm, then, the claim one holds is constituted by one's input and the outcomes which follows are horizontally justifiable based on input differences. Fairness is according to this perspective, cognitively structured on this balance between one's own input/outcome ratio and the input level of some comparable others. In juxtaposition to what was argued in the previous chapter, maintaining a fair balance is not centered upon the idea of reciprocity, rather it is related with maintaining a cognitive balance and, as such, what matters is not maintaining a socioemotional founded social balance, but solely maintaining a cognitive balance, insofar as this cognitive balance is associated with the constitution of one's claim rights as being centered upon maintaining a balance between input and outcome claims – one cannot claim more than one gives. From this perspective it seems to follow that one can obtain claims on other social agents based upon one's investment in them or the relation. Perceived entitlement is, in other words, produced by one's sense of right to claim which is contingent upon how highly one evaluates one's input to be.

Furthermore, the perspective developed here also emphasizes the epistemological dimension of these cognitive claims, in the sense that it argues that these rights are based on the individuals' perceived evaluation which might be biased because one might over-evaluate the value of one's input, just as one's claim rights may also be egocentrically biased. The theory also draws attention to the fact that what enables one to balance one's egoistically biased preferences and moral values is the existence of social norms, thus the *triune* constitution of social rationality is also identified here as the relation between preferences, moral values, and social normative expectations.

### **5.3 Social Utility and Social Preferences**

The modern literature on social preferences is, primarily, based on Ernst Fehr and Klaus Schmidt's (1999) interdependent preference function framework. This entails that before discussing social preferences, I will have to discuss the notion of interdependent social preferences. In traditional utility models, the agent's choice is based on a) her preferences, and b) the utility obtained by the agent is independent of the utility obtained by other social agents. Messick and Sentis (1979), as discussed above, argued

that social agents' choice over alternative distributions depends more upon values than preferences, in the sense that the choice more often reflects a fairness norm than pure preference, and that social agents' choice, when assigning an outcome to Other, reflects not only a fairness norm, as argued by equity theorists, but also preferences implying that social agents hold preferences over the outcome of other social agents. In other words, social agents' preferences concerns a comparative relation between "what they want for themselves" and "what they want for others" – what is most often referred to as social utility. Max Bazerman, Sally B. White, and George Loewenstein (1995) define social utility as follows:

“Social utility functions represent an individual’s level of satisfaction (utility) with an outcome as a function of the payoff to himself or herself and the payoffs to other interdependent parties.” (Bazerman, Loewenstein, and White, 1995: 41)

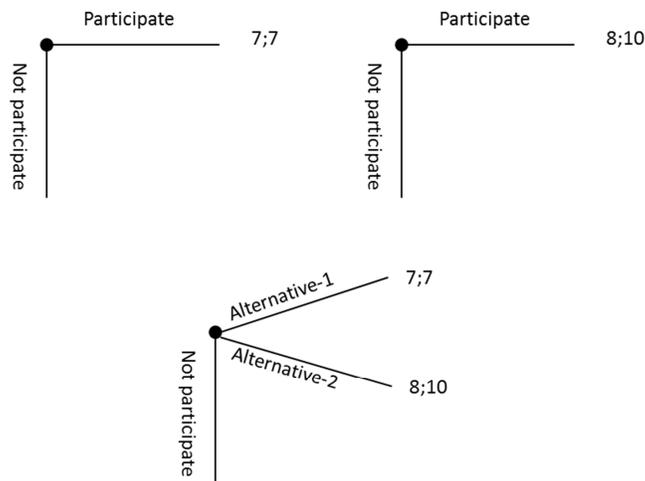
The social utility function is non-linear and its form depends upon different contextual factors, e.g. the quality of the social relation (Bazerman, Loewenstein, and White, 1995: 41). To determine the social utility of each individual respondent who participated in their experiment, Bazerman, Loewenstein, and Thompson (1989: 429, 431) regressed the subjects' satisfaction with each potential outcome, stated in the question, against the monetary value of that outcome to the subject, and the difference between the monetary value of the subjects' outcome and the other person's outcome. A simplified version of the authors' function is showed below with the first component (e.g.  $Outcome_{Person}$ ) suggesting the slope of the function and the squared component suggesting the curvature of the model (Bazerman, Loewenstein, and Thompson, 1989: 428, 429):

$$U_{Person} = Outcome_{Person} + Outcome_{Person}^2 + Absolute\ difference + Absolute\ difference^2$$

Bazerman, Loewenstein, and Thompson's (1989: 438-439) study shows that concerns about interpersonal comparison outweighs concerns about personal outcome. For example, the respondents found an equal split more satisfactory than an unequal split even though the outcomes of both the agents and the social outcome was higher. This suggests that social agents have a tendency to choose a suboptimal outcome for themselves and society due to fairness concerns (Bazerman, Loewenstein, and White, 1995: 41). This tendency was also demonstrated in another set of experiments on independent

choices. In the first set of experiments, the authors found a difference between sequential independent evaluations of choices, and simultaneous multiple comparison of interdependent choices. In sequential choice situations, the respondents' concerns about fairness outweighed their concerns about maximizing their private outcome, whereas in situations allowing multiple choices between interdependent events, the agents choices suggests the opposite – in other words, a preference reversal occurs (Bazerman, Loewenstein, and White, 1992: 229). The results were also replicated in a later study – the choice situations can be illustrated as follows:

Figure 24: Illustration of the difference between sequential and simultaneously weighing of alternatives.



In the first game, the respondents were to choose between participating in a study which paid 7 units to everyone and not participating, and in the second game the respondents were to choose between participating or not in a study which paid themselves 8 units and everyone else 10 units. 72 percent chose to participate in the first, and 52 percent in the second. However, when the two options were evaluated simultaneously, 16 percent chose the first alternative [7;7], 56 percent the second alternative [8;10], and 28 percent chose not to participate (Bazerman, Loewenstein, and White, 1995: 41). This suggests that when evaluating sequentially, fairness concerns outweigh maximization concerns, but not when evaluated together. In other words, as argued above fairness concerns leads to suboptimal choices. It should, nevertheless, be noticed here that fairness concerns only relates to equality and that the respondents demonstrates what Fehr and Schmidt (1999) calls *inequity aversion*. Fairness concerns, here, becomes a bias which

prevent optimization in private life. Fairness has, in other words, lost its moral foundation and has become a bias, see also chapter III. Fairness is, as such, no longer a matter of identifying the middle – that which is fair to me and you – or securing that everyone gets what is justly theirs.

Fehr and Schmidt (1999: 819) states that inequity aversion means that social agents are willing to reject perceived inequitable outcomes – that is, they are, for example, willing to reject the possibility of some material outcome to obtain a more equitable one. The social agents in this model do not just care about their own relative payoff as it would be the case with purely negatively self-centered agents – although the motives are egoistic, because what makes the agent prefer fairness to the other is the emotional discomfort she herself feels in the presence of some disadvantaged other, just as what makes her prefer fairness to herself is the emotional discomfort she feels in the presence of some advantageous other (Fehr and Schmidt, 1999: 822; Camerer, 2003: 103). Fehr and Urs Fischbacher (2002) define it in a later work in the following manner:

“...inequity averse persons want to achieve an equitable distribution of material resources. This means that they are altruistic towards other persons, i.e., they want to increase the other persons’ material payoffs, if the other persons’ material payoffs are below an equitable benchmark, but they feel envy, i.e., they want to decrease the other persons’ payoff, when the payoffs of the others exceed the equitable level.” (Fehr and Fischbacher, 2002: C3)

In that sense, referring to Aristotle’s fairness continuum, what explains the move from “unfair to me” to “fairness” is envy and what explains the move from “unfair to you” to “fairness” is guilt, though as also noted by Colin Camerer (2003: 103) guilt is not accurate, insofar as it requires an action to have been performed. The social agents have different thresholds for advantageous and disadvantageous inequity, as also argued by Adams (1963) above, disadvantageous inequity ( $\alpha_i$ ) weighs more than advantageous inequity ( $\beta_i$ ) on the agent’s decisions – this suggests that negative deviations from some reference outcome count as more than positive deviations, hence the agents are assumed to be *loss averse* in social comparison implying that:  $\alpha_i \geq \beta_i$  (Fehr and Schmidt, 1999: 824). The other-regarding component ( $\beta_i$ ) is defined within the range:  $0 \leq \beta_i < 1$ . The left hand side of inequality, ( $0 \leq \beta_i$ ), suggests that social agents’ dislike being better off than their comparable other in the sense that  $0 > \beta_i$  would suggest that social agents strongly preferred being better off to equity – Robert Franks (1985) has proposed such a

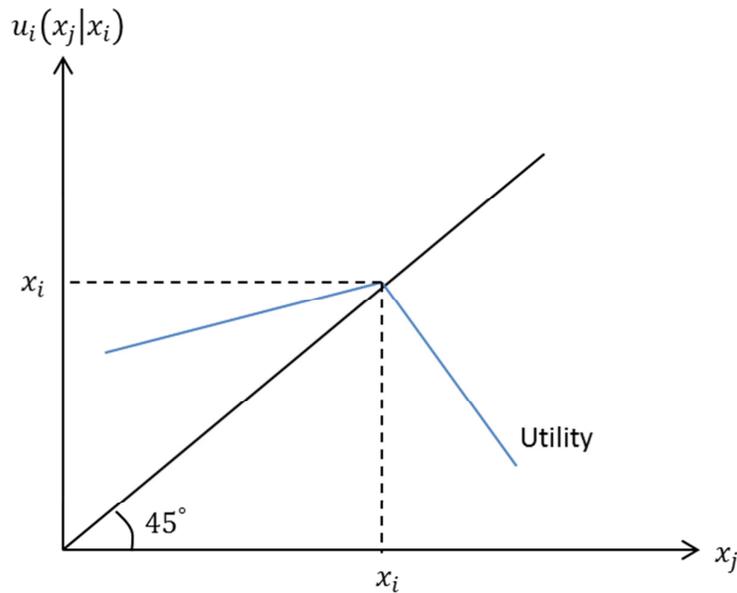
model on negative interdependent utility, stating that social agents are driven by envy or status concerns and, as such, they care more about being the highest paid than being paid their marginal product. A similar perspective is proposed by Barnard (1938/1968: 145) when he argues that pay differences have a symbolic meaning which might overshadow their functional purpose. The right hand side of the inequality, ( $\beta_i < 1$ ), suggests that social agents are not completely altruistic in the sense that  $1 = \beta_i$  would suggest that social agents was willing to throw away a positive material outcome to themselves to offset another social agent's discomfort (Fehr and Schmidt, 1999: 824). Furthermore, insofar as the agents who enter into these games enter in as equals, the referent other must be the other player – an equal – thus the agents are considered to inequality aversive (Fehr and Schmidt, 1999: 822). The simplified utility function can now be written:

$$U_i = x_i - \alpha_i \max\{x_j - x_i; 0\} - \beta_i \max\{x_i - x_j; 0\}$$

(Fehr and Schmidt, 1999: 822)

The social agent's own outcome and the outcome of the other are represented by  $x_i$  and  $x_j$  respectively. The first part of the equation states the utility loss which follows from disadvantageous inequality and the second part states the utility loss which follows from advantageous inequality. Because the agents consider each other as equals, utility maximization occurs at:  $x_i = x_j$  (Fehr and Schmidt, 1999: 822). The utility can be illustrated as follows – as a simplified linear curve:

Figure 25: Utility function with inequality aversion. Source: Fehr and Schmidt (1999: 823).



The  $45^\circ$  line illustrates the proportional relation between the two social agents' outcome, hence in the area above the line  $x_i > x_j$ , and in the area below  $x_i < x_j$ . To illustrate the different thresholds of disadvantageous and advantageous inequality, the slope of the utility curve (the blue line) is rather flat above the  $45^\circ$  line, and below it is rather steep - implying that the marginal effect of disadvantageous inequality is higher (numerically) than the marginal effect of advantageous inequality. The process of comparison which takes place can be illustrated in the following manner: Should the Responder accept/reject the Proposer's offer ( $s$ ) on 50 percent of 100 units? The responder prefers a high payoff to herself and equality between herself and the proposer. If the responder accepts, then, she obtains 50 units and the proposer obtains ( $X_P = 1-s$ ) also 50 units, hence the relational difference between the responder and the proposer is 0 which is the same as if the responder chooses to reject  $\{0,0\}$ , combined with the fact that accepting provides her with 50 units which is more than if she chooses reject - the responder should then accept (Fehr and Schmidt, 1999: 828). Should the responder accept if  $s = 70$  percent? The responder, here, obtains 70 units while the proposer receives 30 units, implying that the relational difference is higher than if the responder chooses reject, however, because ( $\beta_R$ ) has an upper limit the responder will never throw away a material outcome to avoid advantageous inequality - hence the responder should accept. Now consider the opposite situation, should the responder accept/reject if  $s = 40$  percent?

$$\begin{aligned} U_B(\text{accept}) &= sX - \alpha_R[(1-s)X - sX] = sX - \alpha_R[X - 2sX] = sX - \alpha_R X + \alpha_R 2sX \\ &= X[s - \alpha_R(1 - 2s)] > 0 \end{aligned}$$

$$U_B(\text{reject}) = X[s - \alpha_R(1 - 2s)] < 0. \text{ Solving for } s \text{ yields: } s < \frac{\alpha_R}{1 + 2\alpha_R}$$

(Fehr and Schmidt, 1999: 828)

As the inequality implies the outcome does not matter in the last case and insofar as there is no upper limit on  $(\alpha_R)$  nothing prevents the social agent from throwing away material outcome to avoid disadvantageous inequality (Fehr and Schmidt, 1999: 824). If  $(\alpha_R = 3)$  and  $(\beta_R = 0.3)$ , then:

$$\begin{aligned} U_R(\text{accept}) &= 0.4X - 3\max\{0.6X - 0.4X; 0\} - 0.3\max\{0.4X - 0.6X; 0\} \\ &= 0.4X - 3 \cdot 0.2X = X(0.4 - 0.6) = -X. \end{aligned}$$

As the equation above shows, the agent will reject the offer – alternatively this could also have been calculated in the following manner =  $s < (3/[1+(2 \times 3)]) = 0.4 < 0.43$ .

This seems to suggest that social agents dislike inequitable exchanges, insofar as it upsets the initial relational status quo. Relative pay-for-performance initiatives may violate this relational logic, insofar as it might produce social externalities by compensating the better performer higher than the lower performer depending on the relation between the social agents (Bandiera, Barankay, and Rasul: 2005: 948). As argued by Bandiera, Barankay, and Rasul's (2005: 949), if two colleagues consider each other as friends this particular relational structuring might undermine the functionality of a relative compensation system, insofar as the social agents, at some occasions, are willing to internalize the social costs they might impose on one another, because the cost they impose on the other stirs up guilt within themselves – that is, the employees will supply a lower level of effort, if the consequences of choosing a higher one imposes costs on the other. In other words, inequality aversion might cause the agent to produce irrational decisions – not optimal. While this may be considered as irrational from an economic perspective, it might not from a socially rational perspective.

The perspective developed in the second part of this chapter takes into account that outcome fairness is not just based on a cognitive evaluation of the balance between

input and outcome; insofar as it also takes into account the social emotional foundation of relations which emerges from holding interdependent preferences. Furthermore, the perspective also explicates the adverse consequences of fairness when it becomes a goal in itself – when fairness becomes more important than rational optimization. One may, of course, argue that the social optimum is not, necessarily at least, that which produces the highest aggregated outcome to all parties, insofar as this Benthamite assumption does not take into account the future of the social relation which went into the production of this preferred consequence, that is, the distribution which produces the highest outcome to all parties may not be able sustain it in the long run if it is not produced in a social rational manner – in a social sustainable manner. Social rationality as a non-instrumental and non-consequentialistic concept focus, then, on the production of the outcome and from this perspective the socially preferable outcome is not, necessarily, the highest one, rather it is that which is produced in a socially sustainable manner, in the sense that a focus solely on outcomes may not capture the fact that in the long run, if the productive foundation is not maintained, the outcome which from a rational perspective might be preferable may undermine itself due to the social consequences of bringing it about.

This also seems to follow from Bandiera, Barankay, and Rasul's (2005) study, insofar as one might argue that the agents, here, cared about appearing fair – behaving as if they respected the relational bounds between one another – hence, the continuation of the social relation was preferred to monetary outcome. Of course, it might be argued that this concern is not purely moral, in the sense that the agent might simply fear being punished, fear being ostracized from the group, or simply dislike being perceived as a bad colleague. Regardless of the motive, the social rational action has as its functional aim to secure the continuation of social relations. Hence, the distinction between rational optimal option and fairness may reduce the actual complexity of social relations and their meaning – a complexity which is, at least to some extent, captured in the Fehr-Schmidt principle on inequity aversion. Moreover, the perspective developed here seems to capture the necessity of everyday social rationality and not just the necessary kind of social rationality which follows from, for example, teamwork. The effects of a compensation system cannot be understood solely in relation the vertical relation and the traditional focus on fairness, motivation, and satisfaction, insofar as this triadic con-

ceptual relationship does not capture the implications which follows from horizontal relations – cognitively structured as social preferences. The outcome claim may be influenced by one's perceived input, just as it might be influenced by negative social emotions such as envy, but this claim is balanced with a social obligation, socioemotionally founded in guilt and non-consequentialistic concerns about appearing as one who respects the relation by acting fairly towards the other.

#### **5.4 Tentative Conclusions**

The aim of this chapter has been to discuss how fair outcome claims are relationally constituted; vertically constituted in the relation between input and outcome and horizontally constituted as justifiable differences. In accordance with Aristotle's (350BC/2007) definition of the principle of proportionality, the fair is that which maintains the relational status quo. Hence, deviations from this relational status quo are perceived as unfair. The equitable exchange is that which is simultaneously vertical and horizontally fair – that is, it is perceived as a fair exchange according to standard held by the social agent. If the exchange is perceived as inequitable it might cause psychological tension, because the agent might not feel that she is worthy of this in situations of over-compensation, whereas in situations of under-compensation she might not feel that her efforts are recognized and, as such, she might feel inferior to her peers. To restore a cognitive balance between her received compensation and self-perception, the agent might employ different strategies. In the long run, states perceived as inequitable might lead to states of social anomie.

The principle of proportionality is centered upon the perception that social agents care about being fairly compensated, just as it also seems to argue that social agents care about the relative fairness, thus it is accompanied by two central question “is it fair for me?” and “is it fair to the other?”. In that sense, securing fairness in regard to both questions leads to the conception of fairness as the middle. However, as argued by the principle of inequality aversion, the fair may not always be the middle, insofar as it depends on the agent's social preferences, that is, fairness may be, social emotionally, centered between envy and guilt – envy drives the first question, while guilt is the driver of the

other. Allowing perceived fairness to be a matter of social preferences, then, captures the dynamic nature of fairness and its emotional foundation.

This seems to point to the tentative conclusion that horizontal relational concerns might influence the perceived fairness of the compensation system, and this might not be captured within a rational acceptance framework, insofar as it is a matter of social preferences, implying that fairness depends upon the horizontal relational structuring – the relational distributional norms – and their cognitive representation as a sense of envy or guilt. That is, in accordance with the framework presented in chapter III, it might be argued that envy is related with negative deviations from the agent's perceived right, while a positive deviation in the agent's favor is a violation of the rights held by other social agents. Furthermore, while fairness may be viewed in opposition to economic rationality and as producing suboptimal societal results, one needs to take into account that the socially optimal outcome needs not be the rationally optimal, insofar as a long-term perspective on outcome optimality needs to take into account the generative foundation, because if the optimal outcome is produced in a non-social sustainable manner the productive foundation may disappear and, as such, a rational outcome which is not also a socially rational outcome may socially undermine itself when its generative social foundation disintegrates. Hence, the socially rational outcome is that which maintains the social generative foundation. These social rational considerations may also explain why social agents, in some situations, prefer acting fair to maximizing their outcome. This principle is compatible with the Hobbes' notion of the social limits of desire, although the focus here is on individual level; rather the focus is relational, insofar as it is assumed that organizational life is mainly relational in nature – in accordance with Coleman's perception, see chapter III.

## Chapter VI

### Fairness, Cognitive Normative Expectations, and Blame

#### 6.1 Introduction

So far in this thesis I have been discussing the normative foundation of vertical and horizontal relations. In the previous chapter, I argued that outcome claims should be perceived as relationally constituted – vertically centered upon the relation between input and outcome and horizontally justifiable as differences in input. Furthermore, I also argued that the socially rational optimum is the highest achievable outcome which maintains its social productive foundation. That is to say, the horizontal dimension is the productive foundation of social outcomes and an outcome which undermines this ends up undermining itself in the long run.

Cognitively, a fair outcome claim is one that is in accordance with the agent's input, insofar as this proportional relation maintains a cognitive balance, just as it is socially acceptable because it maintains the proportional difference in input. This cognitive balance may also be related with the social emotions, envy and guilt, in the sense that envy pulls towards self-regarding fairness while guilt pulls towards other-regarding fairness – thus, fairness emerges in this balance between what is “fair to me” and what is “fair to you”. The maintenance of this balance is also the foundation of the concept social rationality introduced in chapter III – a concept which emerges from the triadic relation between rationality, morality, and social norms. Social rationality can be viewed as the social dimension of rationality, the dimension which guides social interaction and has as its aim to maintain social relations – vertical as well as horizontal. Reciprocity was, in chapter IV, discussed as a particular kind of social rationality which might govern vertical relations, although it cannot be reduced to this, because vertical relations also rests upon a set of obligations which follows from the formalized nature inherent to the relation.

In chapter III, social rationality was also associated with the responsibility which follows from social relations – from acting into a social world and from being the cause of actions which might have influence on the lives of other social agents. The responsibility, however, associated with bringing about some particular state – for example out-

come distribution – is not directly related with actions produced at the horizontal level as argued in chapters II, III and IV. From a causal perspective, the producer of compensatory differences and differentiation is the compensation system and the personnel who implement its procedures and acts in accordance with its principles. That is, an employee cannot, causally, be blamed for bringing some particular outcome difference about, because the actual producer is, for example, the manager who chooses to reward the particular action. Nevertheless, as argued in chapter II, employees may, at some occasions, blame one another not, however, for bringing about a particular outcome, but for differentiating themselves from the group's norms. Hence, while employees cannot blame one another for bringing some particular state about, they may be able to blame one another for producing certain kinds of actions which are perceived as unfair – a sign of disrespect of the relational foundation – or antisocial. The action performed may, then, be illegitimate while the result may be justifiable. Norms may, in other words, exist on which goals to pursue (legitimate values) and how these should be brought about (legitimate means). Cognitively, norms are related with expectations and, as such, end states which deviate from expectations may have been caused by some abnormal (not normal) event. The agent who differentiated herself by not acting in accordance with the norm, as such, deviated from the expectations and her actions, indirectly, brought about an outcome state different from the expected one. The question is now: can this employee be blamed and is she responsible?

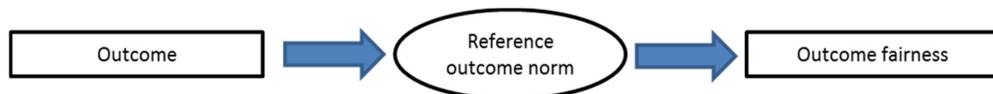
To answer this question, I will take as point of departure Merton's norm theory and Folger and Martin's (1986) theory on relative deprivation, because they introduced the cognitive relation between blame and norms as related with counterfactual reasoning. Folger and Martin's perspective was centered upon Kahneman and Miller's (1986) idea about counterfactual reasoning as the foundation of expectations. Based on this, I will discuss Folger and Cropanzano's (1998; 2001) fairness theory in vertical relations and this I will discuss a possible horizontal extension of this.

## **6.2 Cognitive Normative Expectations and Relative Deprivation**

According to Merton (1938/1968) the "objective" measure to which the individual compares her performance (the causal connection between supplied effort and recog-

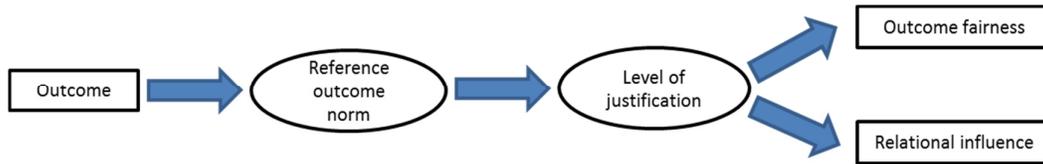
nized effort, see chapter II) is the social norms, just as she from the social norms governing the instrumental dimension, derives an idea about how she *should* pursue her goals. Relational differences can, then, be identified as norms governing particular situations and expressed in the social agents' beliefs on their own entitlement to some object compared to that of member belonging to another social group, and this will have an impact on which goals the agent chooses to pursue, just as the instrumental norms will have an impact on the means with which the pursues it (Merton, 1938/1968: 285). The social norms, in other words, comes to influence with who one compares one's situation, and as such this comparison with some norm becomes a measure of the fairness of the situation, insofar as these becomes *frames of reference* collectively held by a certain group and governing the process of comparison by providing different social categories in which one can place oneself and others (Merton, 1968: 299). As such, relative deprivation understood as standards of evaluation may explain why *objective* states can be differently evaluated by two different social agents, because these bases their standards of evaluation on different frames of reference. The situation, then, has no objective measure of fairness, what matter is the measure applied by the social agent in the particular situation. This may be illustrated as follows:

**Figure 26: Illustration of the causal relation between outcome and outcome fairness with reference norms as social mechanism.**



As the model above illustrates, the fairness of the outcome depends upon the reference outcome norm used by the agent when evaluating her outcome. The impact of the reference outcome norm upon the relational sentimental structuring (positive/negative) is, as argued by Folger and Martin (1986), contingent upon the level of justification supplied by the distributor, e.g. the organization. This may be illustrated as follows:

**Figure 27: Illustration of the causal relation between outcome and relational influence with reference norms and level of justification as social mechanisms.**



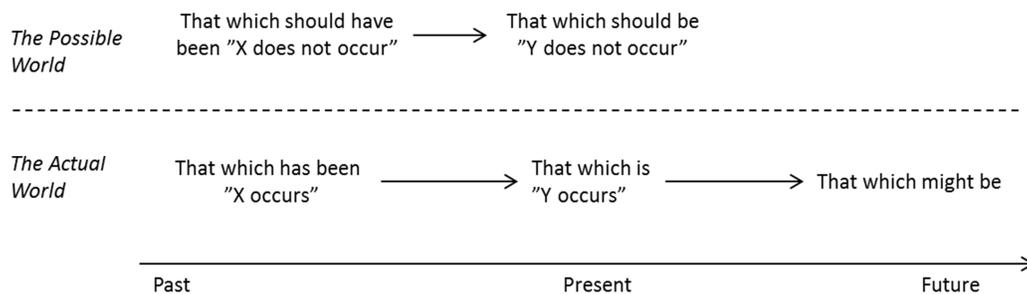
As the model, above, shows, the level of justification not only has an impact upon the perceived outcome fairness, but also on the relational influence of a potential deviation from the applied reference outcome norm (Folger and Martin, 1986: 541). In their study, the authors find that respondents exposed to low levels of justification tend to be more prone to anger and resentment towards the agent who they perceive as being the cause of their situation (Folger and Martin, 1986: 532) compared to respondents exposed to high levels of justification. Furthermore, respondents exposed to information manipulation aimed at providing a high reference outcome norm tend also to be more prone to emotional reactions than respondents with low reference outcome norms (Folger and Martin, 1986: 541). Folger and Martin's (1986) model is built on the idea of counterfactual beliefs – that is, without further justification, the evaluation performed becomes solely based on the outcome reference norms, and as such the actual outcome is compared to the reference outcome, thus deviations from this norm is evaluated as how reality *would* have been. Before I move on, it seems important to fully comprehend the psychological mechanisms implied here, thus I will take a short excursion on the theoretical foundation.

### **6.3 The Foundation of Cognitive Normative Expectations**

Folger and Martin (1986) base their understanding on Kahneman and Tversky's (1973/1979) theory about *availability* and *simulation* heuristics; and Kahneman and Miller's (1986) theory on norms as counterfactual beliefs. Availability heuristics theory is centered on the idea that the frequency or likelihood of an event is contingent upon either the ease/hardship with which something comes to mind (phenomenology of recalling) or the content of the recalled – that is, when judging the frequency with which an event X occurs compared to another event Y, one is likely to base this judgment on the ease with which past occurrence of either X or Y comes to mind and the most fre-

quent event is that which comes easiest to mind; alternatively the content dimension may also influence, in the sense that the number of X's and Y's one can recall generates the foundation of one's judgment and the event which has the highest number of counts/examples is taken to be the most frequent (Kahneman and Tversky, 1973: 164; Schwarz and Vaughn, 2002: 106-107; Kahneman, 2011: 129). The simulation heuristic theory takes a similar perspective, in the sense that it draws on a similar logic of mental associative bonds between likelihood and the phenomenology of recalling. A simulation is the production of an imaginary course of events – or in a counterfactual perspective, it is the production of an imaginary alternative outcome – and the ease with which the alternative course of events is produced and the alternative outcome produced becomes a judgment of how likely the alternative state is to occur (Kahneman and Tversky, 1979: 201-202). Judgments on the likelihood of alternative outcomes is, of course, contingent upon the actual world's degree of mutability – that is, the degree to which it is considered that actual factors could take on alternative states implying that they are either changeable or open to interpretation. In other words, the less open to alternative explanations and the less changeable the real course of events is, the less plausible is the alternative reality (Kahneman and Miller, 1986: 350). Furthermore, the more mutable the course of events seems, the more causal efficacy is attributed to these mutable events, because they *could* have been otherwise which *would* have produced another outcome – see the illustration below.

Figure 28: Real vs. Counterfactual Worlds.



Counterfactual belief formation is a kind of backward processing in which the outcomes of the real world are compared with outcomes of alternative worlds which comes to

mind with some ease and, as such, these alternative outcomes are more or less fragmentary and *ad hoc* in nature in the sense that these are *not* based on mindful (slow thinking would Kahneman (2011) call it) comparisons of similar past events (Kahneman and Miller, 1986: 349). The alternative reality, as such, is partly produced by recalling past events and partly produced by one's imagination (Kahneman and Miller, 1986: 348). The outcomes of the actual reality are compared to the outcomes of this alternative world, however, the aim of this is not to emphasize the difference between how it was (past) and how it is (present); rather, the aim is to emphasize how the outcome of the actual world deviates from the alternative world, in the sense that the plausible alternative world takes on a normative dimension as how it should/could/would be had something been different, and as such it is a variant of the *endowment effect* (Medvec, Madey and Gilovich, 1995: 634) which I will discuss in chapter VII. That is, the hedonic experience of some occurrence is produced by a comparison of the actual with its alternative, and the closer the alternative state seems to be to the actual event, the more intense the pain of an unfavorable comparison and the pleasure of a favorable comparison.

As this perspective, then, seems to argue, outcome fairness is related with the level of justification accompanying it, insofar as the level of justification narrows the gap or reduces the mutability of the perceived course of events. In that sense, by allowing the employee to gain access to the information on which her performance evaluation is centered, the organization might be able to reduce the perceived gap. This might, in other words, be viewed as an early variant of the fair process effect, albeit the focus is on information and not authority.

#### **6.4 Fairness and Blame**

Folger and Cropanzano's (1998; 2001) *Fairness Theory* is based on a similar perspective, although it is more comprehensive in its scope and takes into consideration more than outcome fairness as an interaction between procedural justification and distributive fairness norms. The focus of this theory seems to be the interaction between procedural justification and interactional fairness which is also implied by the subtitle of their paper in which the theory is presented, "*Justice as Accountability*". According to Folger and Cropanzano (2001: 1), fairness theory concerns the assignment of blame, in the sense

that *someone* has *done* something which threatens the well-being – monetary or psychological – of another individual. The fairness judgment, then, can be viewed as a composite of three (theoretically) separate judgments: 1) the present outcome is found unfavorable; 2) The actions which brought about the unfavorable state were intentional (a deliberate choice); and 3) the action which brought about the unfavorable state violates some normative reference standard (Folger and Cropanzano, 1998: 175; 2001: 3). The fairness/unfairness is, now, decided upon by comparing the actual reality with the alternative reality, thus: 1) is the actual outcome more/less unfavorable/favorable than the alternative state *would* have been?; 2) is the individual who brought about the event blameworthy – *could* she have acted otherwise?; and 3) does the actual action deviate from what *should* have been done? (Folger and Cropanzano, 1998: 182,185,188; 2001: 6).

The three modal verbs, emphasized with italics above, signify the counterfactual contrasts, because they explicate the conditional nature of the judgment – “If not X, then Y”. *Would* is conditional in the sense that it compares “what would have been the case” to “what is the case”, and, as such, the degree of aversion can be derived from how *far* the “would-be-state” is from the “actually-is-state” (Folger and Cropanzano, 1998: 182; 2001: 11). Nicklin, Greenbaum, McNall, Folger, and Williams (2011: 133) also found that the strength of the experience of unfairness was related with this perceived degree of distance between the actual and alternative outcome. *Could* is conditional on *can* implying that the alternative set of actions considered must be feasible, meaning within the space of the possible actions of the agent (Folger and Cropanzano, 1998: 185; 2001: 11, 18). Experimental studies have demonstrated that contextual factors such as expertise and knowledge have an influence on the production of counterfactual thoughts and the strength of the unfairness experience (Nicklin, Greenbaum, McNall, Folger, and Williams, 2011; Nicklin, 2013). Mistakes made by, for example, experts and knowledgeable agents are more likely to produce counterfactual thoughts, just as the strength of the experience of unfairness is higher than in cases when these descriptions are not attributed to the harm doer (Nicklin, Greenbaum, McNall, Folger, and Williams, 2011: 131, 137; Nicklin, 2013: 52). Nicklin (2013: 56) demonstrates the opposite effect that non-mistake actions made by experts are less likely to produce counterfactual thoughts

than non-experts, hence social agents are more likely to accept (perceive as fair) unfavorable actions made by experts.

The other person who becomes the target of these could-judgments is the person whose actions, in some way, either caused the event to occur or preceded the event in a manner which eventually led to the present state – that is, a person whose actions either directly or indirectly caused the present state to occur. In accordance with Kahneman and Tversky's (1979) simulation heuristic, the easier an alternative action comes to mind, the more blameworthy is the person (Folger and Cropanzano, 2001: 18). Moreover, besides the causal-link, the actions of the agent, who is target of this kind of judgment, must also be the result of an intentional choice – that is, actions which are the result of a deliberate choice made by the agent, unintentional consequences and acting on orders, however, are not valid exceptions because implicit in the judgment is an assumption on moral agency (Folger and Cropanzano, 2001: 12). Lastly, *Should* is the conditional form of *shall* implying that the deviation is a normative deviation from a “how-it-should-have-been” state or more accurately a “how-it-shall-be” state. This should-judgment, designates that a norm violation have occurred in the sense that it deviates from how it shall be, the normative frame of reference (Folger and Cropanzano, 1998: 188; 2001: 21). The abnormal action, then, becomes the norm violating action, insofar as alternatives to this are easily available, whereas the normal action is in accordance with the norm, and as such it is similar to the actual event. Hence, the action which violates the ethical and moral reference norms is an abnormal action within a particular context (Kahneman, 2011: 73).

In that sense, unfairness requires the violation of some ethical or moral standard governing how social agents should act – this latter point is also in accordance with Mill (1863/2001: 51) who in his essay “*Utilitarianism*” argues that the idea of judgment presupposes a rule of conduct – a rule which if violated produces this sense of unfairness. Mill (1863/2001) adds that the rule is accompanied by a distinct emotional foundation, which Folger and Cropanzano (2001) does not consider because the foundational cause of this sense of unfairness, here, is backward reasoning and not the invoking of some emotional sense of unfairness. Folger and Cropanzano's (2001) idea of fairness seems to resemble Aristotle's (350BC/2007) idea of corrective or retributive fairness, which I

identified in the previous chapter. Aristotle (350BC/2007: V.8, 1135a-1135b/137-138) also emphasized the intentional dimension of the *harm doer's* action when he distinguishes between actions which emerge from the free will and those that do not. The action which springs from the free will can be a deliberately planned action, it is, nevertheless, chosen by the agent (meaning that it is within her power to choose otherwise) and it is characterized by the agent's consideration of the action's consequences (the means which it implies and the target) meaning that the action as well as its consequence is intentional. The action which does not spring from a deliberate choice or which has unintentional (not foreseeable, illogical or beyond common sense) consequences can be treated as wrong, but not unfair. In other words, according to Aristotle (350BC/2007) the attribution of unfairness requires the attribution of agency and intentionality.

Folger and Cropanzano (2001), then, add to this framework that intentionality and agency is vulnerable to mental heuristics, insofar as it matters how easy the alternative comes to mind, thus the more easy an alternative comes to mind, the more blameworthy the person seems to be. Besides the attribution of blameworthiness, this also seems to attribute causal power, in the sense that the more easy an alternative comes to mind, the more easy it is to believe that the blameworthy person's actions also were a necessary condition for bringing the event about, hence had she acted otherwise this would not have occurred. Kahneman and Miller (1986) also finds a strong connection between "close calls or near misses" and possible actions – that is, the closer the agent is to achieve a goal, the more she will blame herself that she did not make some alternative action, the action that she did take, then, is seen to have a great causal power, in the sense that it, alone, brought about the actual outcome. The connection between the actual outcome and the outcome, as well as the connection between the alternative outcome and the alternative, must, then, be strong causally connected (a normally occurring event under such and such circumstances) in a frequently recurring pattern, in the sense that this particular action *always* leads to this particular outcome. The attribution of blameworthiness, then, implies 1) the attribution of intentionality and intrinsic causality (agency); and 2) the attribution of external causality – the actions of the agent actually brought about the event. Taken together, these two premises requires *causal agency* to be present.

From a vertical perspective, then, it makes sense to discuss the degree of blameworthiness as contingent upon the factors just mentioned, insofar as direct causality is more or less assumed. In that sense, the question is not so much whether or not the actual state is a result of the blamed agent's actions; rather the question is a matter of the degree of blameworthiness (Moore, 2009: 21). From a horizontal perspective, however, one cannot establish the same assumption on causality, insofar as the end state seldom follows solely from the actions of the receiving agent, but implies the mediating role of a reward system and some agent who acts out its principles and this agent is (see section 4.1), in the terminology applied by Folger and Cropanzano (1998; 2001), the direct causal agent. Notwithstanding this, if causality cannot be established, the question becomes a matter of the relation between causality and morality, because if responsibility is related with, as argued by Aristotle, intentionality and being-the-cause, as argued by Kant, then the employee who brings about the state may not be blameworthy, inasmuch as no direct causal relation can be established, just as the end state may not be assumed to be intentionally brought about, as such, it is more to resemble with a by-product, thus the agent is not the cause and, as such, the question is not a matter of the degree of blameworthiness; rather, the question becomes a matter of blame-*ability* – whether the individual can be blamed at all, just as it becomes a matter of the relation between causality and morality – that is, to which degree is one responsible for bringing about states which are purely by-products if at all?

In philosophy of law, the question of the relation between moral responsibility and causality is often related with the relation between intention (exposing to risk) and caused-outcome (actualized intention – caused harm), in the sense that an agent who intends to bring some state about is just as blameworthy as if she had brought it about, because responsibility is something which one *earns* through intending – that is, since one cannot control whether some harm is actually inflicted upon another agent in accordance with one's intention, one is simply responsible for exposing the other to the risk of this harm being realized – if this was not the case, law would be centered upon a conception of responsibility in which moral responsibility was simply a matter of luck (Moore, 2009: 22; 25; Wolf, 2013: 7). From this it seems to follow that an agent is not responsible for bringing the actual state about; rather, he is responsible for intending and acting on this intention – moral responsibility, from this perspective then, is not

associated with causality. Here, of course, causality is a matter of control, in the sense that it is a matter of what is within the agent's control, on the one hand, it might be argued that the agent is solely responsible for his action because what happens as result is beyond his control, whereas, on the other hand, it might also be argued that the agent is responsible because he controls something besides himself, although he is not in complete control of the result – related with the *battle* between *incompatibilism* and *compatibilism* in philosophy of law on the notion of free will (Moore, 2009: 26). In order to associate responsibility with causality, Moore (2009: 30; 31) argues that what matters is not so much this intellectual battle as it is first-person experience – that is, an agent who brings some end about which harms another agent feels guilt, he feels guilty because *his* actions led to the harm of another individual – he does not regret having performed the action, rather he feels guilty for having imposed the harm and, as such, guilt is related with having caused, or believe that one have caused, harm. In accordance with this perspective, the agent who causes harm is more blameworthy than the agent who solely intends to do it or who risks doing it (Moore, 2009: 32). In that sense, in matters associated with guilt, perceived causality matters more than issues related with free will. Furthermore, from this first-person perspective, another distinction seems to emerge between causal-responsibility and non-causal-responsibility – a distinction between being a direct cause and the agent *merely* believing that the state would not have occurred had he not acted in a particular manner (Rosen, 2011: 408). This is, of course, a matter of the underlying assumptions on the metaphysics of causality, a matter which is closely related with not the cause of the action, but the causal relation between the action and its result, in the sense that in order to understand matters related with indirect causality, one must take into account the mediating events which eventually leads to the cause (Rosen, 2011: 414). Moore (2009), however, refuses such a reductionist distinction between cause and non-cause when he introduces a distinction between direct-causes and near-causes – an indirect causal relation is a constituted relation and the cause follows from this and not from the actual course events – thus, whether or not one's actions or omissions directly caused the result is of less importance than the question of whether or not a cognitive relation can be established (Rosen, 2011: 415). One might also distinguish between causality related with responsibility and causality related with results – in the sense that one may distinguish between the act of taking a gun and shooting another

individual (causality of responsibility) and matters related with how the gunshot actually killed the other individual (causality of results) (Rosen, 2011: 413).

The matter discussed here is not normative in character and related with whether or not one *should* be blamed; rather the matter here is simply a matter of reflection. Hence, in what relation stands the moral responsibility discussed in chapter III to the causality of compensatory differences? From what has been argued, so far in this section, one might argue that a qualitative difference exists between actually causing the result and the result simply following as a by-product, and that results which follows from as a matter of indirect causality depends on the establishment of a cognitive causal relations. The problem, here, however is not whether the result followed from an actual intended action, omission, negligence, or fault; rather the central question, here, whether a similar line of reasoning applies when the result is not intended, but follows as a pure by-product, insofar as it is assumed that the agent who brings about the outcome state is not trying, directly at least, to bring about a particular compensatory difference, rather she is solely pursuing some pre-established goal or an opportunity of being rewarded. Implicit in this, then, is the assumption that status ranking and envy are not causes of results, but solely causes of actions, although this distinction might be less useful if, as argued above, the agent cares more about being the highest paid than maximizing her outcome. If envy and guilt are balanced with one another, as argued in the previous chapter, it must, of course, be assumed that the agent holds social preferences over which outcome states she wants to bring about and, as such, she might try to bring this about. This outcome state, nevertheless, cannot be causally related with the actions of the individual social agent, insofar as it depends upon other external factors which are not within the control of the agent. The agent, then, is not in control, completely at least, of the results of her actions. In comparison with the analogy on the difference between shooting and killing, one might question if this distinction also applies if it implies the intervention of other agents as it does in the case of compensation. From the discussions briefly introduced above, it seems to follow that at the very least one must presume a relation between the harm-doer and the harmed, in the sense that some causal relation, direct or indirect must be established. Notwithstanding this, as argued by Moore (2009) first-person experience is less concerned with establishing metaphysical relations, hence a social agent may still feel guilty in indirectly bringing about some state regardless of

whether or not she is actually causally responsible. Moore (2009) would not, as demonstrated above, accept this account or at least he would argue that the qualitative sense of guilt – the intensity of the emotion is very different in the two cases. Wolf (2013: 14), on the other hand, introduces what she calls the *nameless virtue of responsibility* to explain why social agents in many cases feel guilt even when there is no direct causal relations, because, as she argues, feelings of guilt goes beyond matters related with the result of some action or regret associated with performing a particular action. She defines this nameless virtue in the following manner:

“It involves living with an expectation and a willingness to be held accountable for what one does, understanding the scope of “what one does,” particularly when costs are involved, in an expansive rather than a narrow way. It is the virtue that would lead one to offer to pay for the vase that one broke even if one’s fault in the incident was uncertain; the virtue that would lead one to apologize, rather than get defensive, if one unwittingly offended someone or hurt him.” (Wolf, 2013: 13)

From a first-person perspective, then, one might hold oneself responsible for results which may solely be indirectly causally related with one’s actions or perhaps not at all. The agent might, in other words, feel guilty of bringing about some state which harms another agent – make her worse off in some manner. As I will argue in chapter VIII, the feeling of guilt is associated with moral transgressions of feeling like one has wronged another human being – acted against one’s moral obligation towards this other. This sense of obligation is, at least to some extent, a matter of the moral values one holds, just as it might also be influenced by the social norms governing particular relations, insofar as argued in chapter II, the agents may be allowed to bring about a particular outcome difference, but they may not be allowed to perform the actions necessary for bringing it about, because this would come to symbolize a norm transgression. That is, the emotional foundation, here, may not be regret and guilt, but guilt associated with having imposed harm on another agent and shame from having transgressed the governing social norm. Hence, from a third-order perspective, the agent may not be blamable and this might be why norms are not produced which prevent agents from bringing about some particular states, rather norms govern particular actions – from a third-order perspective, the agent is not responsible for bringing about some particular result, but for producing a particular kind of actions (norms are non-consequentialistic in nature, as

I will argue in chapter VIII) regardless of which consequences it might have. From a first-person perspective, however, this distinction may be of little importance, in the sense that this might not influence how one's feels.

### **6.5 Tentative Conclusions**

The aim of this chapter has been to discuss the normative cognitive constitution of the outcome claim. In the previous chapter, I discussed how this claim may emerge horizontally from differences in input and vertically as a matter of input – hence an unfair outcome was one which did not main a horizontal sense of proportionality as differences in input and vertically, one which did not respect the proportional relation between input and outcome. In this chapter, I have discussed the cognitive normative foundation of deviations from outcome expectations – when the expected outcome deviates from the expected one. To secure a theoretical relation between the previous and the present chapter, the theoretical point of departure was Merton's conceptions of social norms and Folger and Martin's theory on social deprivation and its relation with information. In accordance with Merton's conception of social norms, the agent might pursue particular goals based on particular means and in opposition to the rational perspective, the agent is not pursuing the goal because she desires to bring a particular state about, rather she is pursuing socially legitimate goals. Claim rights may, then, according to this perspective, emerge from normative expectations and deviations from these may cause psychological anxiety. Folger and Martin argue, however, that this psychological anxiety may be reduced if the agent is informed about why this deviation occurs, because it limits the agent's possibility of producing counterfactual beliefs.

Fairness may, in accordance with this, emerge as a result of counterfactual beliefs – as when the reward one receives deviates from one's expectations about what one should have received had something been different. Hence, fairness becomes a matter of assigning blame. From a vertical perspective, the actions of the manager may have directly caused that the actual situation deviates from the expected one and, as such, her degree of blameworthiness must be determined centered upon the nature of deviation, whether she could have acted otherwise, whether she is actually accountable, etc. While this discussion makes sense from a vertical perspective, it might be more problematic to

apply horizontally, insofar as the agent who brings about some particular end state may not be directly accountable and neither may she want to bring this state about – thus, the state may be a pure by-product.

From a third-person perspective, the employee may not be causally responsible for bringing this state about and, as such, she is not blamable – norms may, nevertheless, be imposed preventing the agent from conducting the action regardless of its potential consequences. From a first-person perspective, the distinction between being the cause or not may not prevent the agent from feeling guilty – feeling as if she has transgressed some moral norm by inflicting harm on another social agent. This feeling of guilt may, in some cases, be accompanied by a feeling of shame when the agent also performed an action which she perceived as violating a governing social norm. Hence, while the relation between moral responsibility and causality is problematic when one shifts from the vertical to the horizontal relational level, it seems that one might either associate it with a moral responsibility which follows from performing the action or a moral responsibility following simply from outcomes which may in some sense, perhaps only vaguely, be associated with agency. Social rationality, from a causal perspective, cannot emerge from bringing about some particular end state although it may indirectly be associated with the harm this imposes on other agents. This harm, however, cannot be directly associated with one's agency, in the sense that one does not decide whether one's actions are rewarded or not. Social rationality cannot, then, be centered on causal determination of responsibility on consequentialistic terms, but must be non-consequentialistic and related with the production of actions – not outcomes.

## Chapter VII

### Fairness and the Hedonic Status Quo

#### 7.1 Introduction

In the previous chapters, I have discussed two dimensions of the cognitive foundation of outcome claims. In chapter V, I discussed how the claim could be founded on input vertically as well as horizontally. In chapter VI, I discussed how cognitive normative claims emerged from comparisons with counterfactual states might also be part of cognitive foundation of claims. Furthermore, in the previous chapter, I also discussed how the relation between causality and morality influenced perceptions on fairness in vertical as well as horizontal relations. From the cognitive constitution of the claim, discussed so far, also follows a particular constitution of fairness, insofar as it in chapter V was argued that the claim is vertically fair if proportionality is secured in the relation between input and outcome, just as it is vertically fair if it maintains horizontal proportionality. In chapter VI, it was argued that fairness and unfairness emerges as matters of comparison between the actual world and the potential world, and if the actual world deviates from the potential world – the world which should have been actualized – someone must be blamed, in a corrective sense, the blameworthy agent whose actions caused the deviation to occur must be identified and her blameworthiness established. The cognitive normative foundation of fairness is, in other words, centered upon an assumption about a fair counterfactual world and that deviations from this are abnormal and unfair. Common to the cognitive foundation of both models, then, is the assumption on comparison – social comparison in the model based on proportional fairness and possible world comparisons in the other. Unfairness emerges in deviations whereas fairness emerges in situations when the world is as one expects and one's relational stand is maintained.

In the present chapter, I will discuss a third cognitive dimension which goes into the constitution of the outcome claim – temporality. In chapters III and IV, I argued that horizontal as well as vertical ties bind; hence social rationality was introduced as a concept which explained how social agents navigated within these social relations. In chapter V, it was argued that relational proportionality binds, just as it was argued, in chapter VI, that one's past experience binds one's present expectations. In the present chapter, a

third kind of tie that binds will be introduced – temporality and hedonic adaption will be viewed as cognitive ties which binds or influences how one perceive and evaluate the present situation. Fairness and unfairness emerges, here, as a matter of comparison between the past and the present state expressed in terms of hedonic experience.

The imbalance to which Aristotle (350BC/2007: V.4, 1132b/131) was referring in chapter V was not just between two individuals, insofar as it can be extended to refer to an imbalance between two states, *the before* and *the after*. *Profit*, he argues, is when the after is larger than the before (before < after), the opposite, *loss*, is when the before is larger than the after (before > after) (Aristotle, 350BC/2007: V.4, 1132a/130). The just, then, is when the before equals the after (before = after). Note, however, the individual who gains a profit has done so at the expense of another whom, logically then, suffered a loss, fairness, as such, has character of a zero sum game. Kahneman, Jack L. Knetsch, and Richard H. Thaler (1986b) argue on this matter in their cognitive approach to fairness:

“The cardinal rule of fair behavior is surely that one person should not achieve a gain by simply imposing an equivalent loss on another.” (Kahneman, Knetsch, and Thaler, 1986b: 321)

The loss experienced by the aggrieved social agent may not only refer to a material loss, but also to the *pain* which the actions of another social agent have brought upon her by conducting certain actions. In that sense, loss may refer to the hedonic experience associated with such a loss. I have already in chapter III presented the concepts of pain and pleasure as the hedonic foundation of actions.

## **7.2 Morality as Irrationality**

I initiated this chapter with an introduction to Aristotle’s idea about the temporal dimension of fairness which stated that gains and losses are measured as differences between two temporal entities, the before and the after – the past and the present – and as it is the case with his understanding of fairness, status quo, when past equals present was considered as fair. This particular conception of fairness is also applied by Kahneman, Knetsch, and Thaler (1986a; 1986b). In the previous chapters, I have already discussed the emotional stress which follows relative deprivation and inequitable relations, just as

I also discussed how equity may be considered as a relational norm and that it was deviations which caused this emotional stress. The social preference tradition added to this discussion that judgments on inequity could just as well be reflections of envy and guilt. The point of departure in the previous discussion, however, was relational concerns and, as such, the temporal dimension was left out. Kahneman, Knetsch, and Thaler (1986a; 1986b) then add the temporal dimension, but leave aside the relational dimension at least to some extent. The perspective developed in this chapter, then, may be called the cognitive approach to fairness, because it encompasses not just the motivational dimension of fairness, self-interested versus non-self-interested motivation, but also the hedonic and epistemological dimension of fairness. Elster (1992) demonstrates, in accordance with this tradition that different principles of fairness are chosen by social agents to solve particular problems and as such the choice of a fairness model is based on (1) beliefs about fairness, and (2) preferences on different end states often regardless of pragmatic concerns. For example, the fairness principles in the allocation of rewards may be seniority or status related to biophysical properties (i.e. gender, ethnicity etc.) (Elster, 1992: 74, 76).

The cognitive approach to fairness, takes as point of departure the experience of unfairness or perhaps more specifically the *psychophysical* sensation of unfairness, insofar as they push past the dimension of cognitive uneasiness and discusses it terms of pain and pleasure as related to deviations from status quo; and the epistemological dimension related with the establishment of this status quo. The epistemological dimension emphasizes the fact that judgments on fairness/unfairness are vulnerable to framing effects, that is:

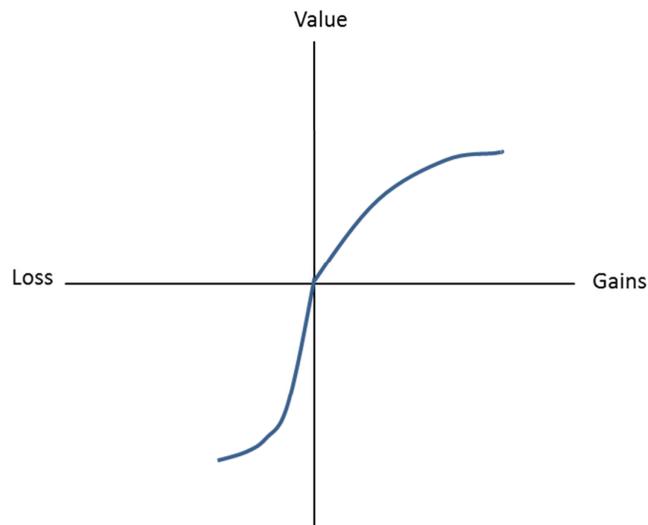
“The determination of the reference level in a choice is subject to framing effects, which can yield inconsistent preferences for the same objective consequences. Similarly, judgments of fairness cannot be understood without considering the factors that determine the selection of a reference transaction.” (Kahneman, Knetsch, and Thaler, 1986a: S297).

The concept of framing applied here is, to some extent, similar to the one proposed above, although here it does not have to be relational and it is not based on social comparison, implying that in the framework proposed here, it is purely vertical, in the sense that it solely concerns the relation between the organization and the employee. It seems, however, rather unproblematic to add the horizontal dimension to the model which also

seems to what Fehr and Schmidt (1999) are doing when they distinguish between the relational difference between before and after, see my discussion of inequity aversion in chapter V.

The implications of the choice of reference point is associated with the two phases of choice that Kahneman and Tversky (1979: 28, 30) identifies in their “Prospect Theory” – the first phase of choice is framing or the coding as gains and loss relative to some neutral reference point; and the second phase of choice is the evaluation of each possible choice against one another to identify the one offering the highest level of utility. From an economic theoretical perspective, the existence and implications of cognitive frames violates the principle of invariance insofar as the choice itself not only depends upon probability weighed consequences, but is vulnerable to framing effects, that is, these cannot be taken as exogenously given and objective because the probabilities as well as the payoffs are contingent on how it the prospect (risky asset) is framed (Kahneman and Tversky, 1984: 4; see also chapter III). The main consequence of this with regard to choice is that the value of the asset is subjectively determined relative to the asset’s position (the reference point) and the magnitude of the change (positive/negative) relative to the previous established reference point (Kahneman and Tversky, 1979: 32). This might be illustrated as follows:

Figure 29: Value function (Kahneman and Tversky, 1984: 3).



As this graph illustrates, the value of the outcome depends on whether it is considered as a loss or a gain relative to the point of reference point (origo), and gains are framed as positive or favorable compared to losses which are framed as negative and unfavorable (Kahneman, 2011: 280). The curvature of the graph illustrates two important matters: (1) diminishing sensitivity; and (2) the asymmetry of gains and losses, because at the concave part of the curve (in the gains area) the slope is less steep (in numerical terms) than with regard to the convex part (the loss area) implying that the positive effect of a gain may be outweighed compared with a similar negative effect of a loss – or for short, people are more sensitive to losses than to gains, hence they are *loss averse* (Kahneman and Tversky, 1984: 3; Kahneman, 2011: 282). This entails that the prospect of a possible loss outweighs the prospect of a possible gain making agents act in a risk averse manner with regard to gains; and risk seeking with regard to losses, because the negative effect of loss is so high (Kahneman and Tversky, 1984: 3). Consider, for example, the prospect of losing 100 with certainty and 110 with 85 percent probability – because of the 15 percent probability of losing only  $(110 - [110 * 0.15])$  93.5 units, agents often choose the certain loss over the gamble. Furthermore, prospect theory also demonstrates that decision weights are not equal to the probability of the event, because agents tend to treat low probability events as possible (amplifies the aversiveness of severe, but unlikely losses making agents act risk averse; and enhances the value of improbable gains making agents act risk seeking); and high probability events as certain (the high probability of loss makes agents act risk averse, whereas the high probability of gains makes agents act risk seeking)(Kahneman and Tversky, 1979: 37; 1984: 7; Kahneman, 2011: 317). The high probability event which appears certain presents itself as an expectation on a certain future loss or gain, hence even the slightest chance of loss turns the expectation on a future gain into a fear of disappointment, just as even the slightest chance of avoiding a loss turns the expectation on a future loss into a hope on avoiding it.<sup>34</sup> In a similar vein, even the slightest possibility of a high outcome invokes hope, whereas even the slightest possibility of a large loss turns hope into a fear of a large loss (Kahneman, 2011: 317).

---

<sup>34</sup> In some cases it may be argued that when highly probable losses turns into certain losses they will have a counterproductive effect in the sense that they invoke despair in the agent – that is, when all hope for a better future is extinguished, fear turns hope into despair (Elster, 1999: 82).

From what has been argued so far, the cardinal rule of fairness is that one agent is not allowed to impose a loss on another agent. The loss is determined as the negative deviation from some cognitively established reference point. The negative psychophysical experience of loss is greater than the positive expectation of gains; hence if the unfair deviation is experienced as loss, it must be associated with a negative or painful hedonic experience. What is established here can be associated with decision heuristics or cognitive rules applied in order to produce a response on a particular manner and, as such, these rules represent psychological mechanisms explaining the relation between text, context, and choice. In that sense, a particular choice is evaluated within a particular decision-context and the value on which the assessment is based is relative to the point of reference. This entails that choices between different prospects as well as judgments on fairness are vulnerable to framing effects – in fact they emerge from these, implying that the agent passing the judgment or making the choice have no reason to do this, insofar as this occurs more or less automatically as result of these cognitive mechanisms associated with retrieving the necessary information from the context, encoding this as either a loss or gain, and passing the actual judgment in accordance with some rule. From an economic rational perspective, the decision chosen and the very process of choice may be viewed as irrational. The irrational aspect of the perspective may, of course, also associated with the process of choice and the information which goes into this, in the sense that the choice is not based on careful and correct assessments; rather, it is centered upon a reduction of complexity in which the mind seeks the easiest way to a particular judgment. The perspective described is, nevertheless, descriptive in nature; hence the authors did not try to propose a normative theory. In accordance with what was argued in chapter III, one could argue that reducing choice or judgment simply to a matter of irrationality may represent a simplification, insofar as to assume that these judgments or choices are made without being based on any reasons seems to neglect the rational competence of agents. The argument proposed, here, is not that rationally based decisions or judgments are more correct in its production, because the agent may retrieve the right information and draw the right inferences without passing a correct judgment, just as she might pass a correct judgment based on the wrong information and inferences. That is, what is argued here is, simply, that one should not neglect that social agents might have reasons – albeit these may be wrong – just as one

should neither equate incorrect judgments with irrationality and lack reason nor rationality with the opposite (Boudon, 2001: 132). Notwithstanding this critique, enquiring into the cognitive foundation of outcome claims may enable an understanding of how these claims are constituted and also how they come to be inter-subjectively justifiable horizontally as well as vertically.

### **7.3 Hedonic Experience as the Foundation of Fairness**

The reference point, then, determines the subjective value which the prospect takes on, just as it influences the fairness judgment by generating the foundation on which this is centered. The reference point, in other words, is the *before* to which Aristotle (350BC/2007) referred above. In exchange situations, the before refers to the past history of similar transactions which have occurred between the parties unless these are explicitly referred to as temporary (Kahneman, Knetsch, and Thaler, 1986b: 319). That is, in an employment relation, for example, the organization is not allowed to lower the employees' salary if its business is going well, but unemployment in the area has increased (Kahneman, Knetsch, and Thaler, 1986b: 320). It is, however, acceptable that an organization protects its profit at the expense of its exchange partners, just as it is acceptable if the organization hires a new employee to a lower salary or lowering the salary of the employee if she takes on a new job within the organization (Kahneman, Knetsch, and Thaler, 1986b: 319, 320). Stated simply, an agent is not allowed to impose losses on another agent, unless this agent's own reference level is threatened. As this seems to explicate, in exchange situations both parties are entitled to their reference level. The authors add to this:

“It should perhaps be emphasized that the reference transaction provides a basis for fairness judgments because it is normal, not necessarily because it is just.” (Kahneman, Knetsch, and Thaler, 1986b: 321)

In this quotation *normal* is used in a similar way as Kahneman and Miller (1986) did above, as that which is not abnormal, see chapter VI. The reference point may, then, be understood as being entitlements or expectations produced by a past history instituting a sense of normality through the workings of adaptive preferences (Kahneman, 2011: 282).

This matter becomes important when the analytical view shifts from third person judgments to first person experience – that is, the conceptions of fairness developed so far in this chapter has been based on survey results in which the respondents were to judge the fairness of an exchange between two parties, the perspective in this thesis, however, is on first-person experiences. In accordance with Kahneman and Miller’s (1986) theory on normal events, the reference point takes on a particular dimension because it comes to be a remembered “before” to which the present experience is compared, thus as already argued deviations becomes deviations from how it *should* have been. This remembered normal event then generates expectations, which might either be confirmed (the normal event) or deviate in a negative sense and invoke regret, envy, disappointment within the agent; or in a positive sense and invoke relief (Kahneman, 1999: 12; Elster, 2007: 150). Furthermore, these comparisons between “how it should have been” and “how it is” has a tendency to intensify the emotional experience felt by the agent – that is, the experience of the positive deviation is intensified, though due to hedonic adaption the intensity of the experience decreases (but does not disappear) when the positive deviation becomes the *new* standard of comparison (Kahneman, 1999: 13; 2000a: 686). One should, nevertheless, be careful when comparing this to the value function presented above, because (1) the neutral reference point does not necessarily represent a hedonic zero in the sense that it represents a hedonic adaptive level; and (2) one cannot necessarily interpret loss aversion in hedonic terms as the desire to avoid pain, because loss averse behavior may also be interpreted as ingrained conservatism (Kahneman, 1999: 18).

Now, as discussed above, one should distinguish between remembered utility and present utility, insofar as the former belongs to the past while the latter is centered in the present moment. The present moment and the experience which accompanies it exists only in the *now* – the present when they are experienced – and when the now has passed into a *after* the experience of that was is irreversible – it cannot be changed – that is, pleasant memories of an unpleasant period does not, at least from a rational perspective, change the qualitative nature of that moment (Kahneman, 2011: 381). The difference between now and before emerges, because the total experience of the event is not equal to an integration of the moment-by-moment accounts; rather the retrospective evaluation of an event is based on what Kahneman (1999: 19; 2000b: 697; 2011: 384) calls the

*peak-end* rule – that is, the hedonic experience of an event is the average of the two events, the peak experience and the end experience. Additionally, the shorter the time-span is between these two events, the more intense the experience, because hedonic adaption has not occurred, thus by allowing hedonic experience to occur, the intensity of the experience decreases and the remembered utility may as a consequence present itself as more pleasurable or painful than it actually was – hence, remembered utility might make the agent choose the choice which presents itself with the least unfavorable utility as it was the case with the subjects in one experiment when they were to choose between immersing their hands into cold water for a short or long time period, the majority chose the second alternative, because the memory of this was less unfavorable than the first due to hedonic adaption related with the length of the time-span between peak and end (Kahneman, 1999: 20; 2011: 381, 382). Kahneman (2011) describes the difference and its implications with his usual clarity:

“The experiencing self does not have a voice. The remembering self is sometimes wrong, but is the one that keeps score and governs what we learn from living, and it is the one that makes decisions. What we learn from the past is to maximize the qualities of our future memories, not necessarily of our future experience.” (Kahneman, 2011: 381)

Kahneman (2011) seems, here, to be arguing that agents cannot necessarily trust their preferences to reflect their best interest, because the evaluative process and eventual choice is based on remembered utility which might very well be wrong. In other words, remembered past hedonic experience turns into a particular hedonic experience which in certain situations is recalled as the base of comparison, just as it seems to have a direct influence on the agents' choice because the remembered hedonic experience comes to be reflected in the present choice, and as such it shapes the preferences of the agent. The agents' preferences, in other words, does not represent what the agents believe will bring them the highest level in the future; rather what has brought them the highest level of utility in the past. The agents adapt their preferences to past hedonic experience. In chapter III it was demonstrated that the classic philosophers focusing on preferences, either centered their idea on preferences on reason (Plato and Kant) or on hedonic experience (Hobbes, Hume, and Bentham). Insofar as Kahneman's hedonic psychology can be considered as a restatement of Benthamite hedonic philosophy, the argument just discussed complicates the matters of choice, because this discussion extends the classic

discussion on the epistemological problems associated with hedonic experience. The question, here, is not whether that which feels pleasurable is actually good in an evaluative sense; rather, the question becomes a matter of whether or not one can actually trust this hedonic experience as being in accordance with one's interests – in some sense then, this mistrust of one's experience can be associated with the Hobbesian argument on the primary principle of rationality – what matters is not whether something is in accordance with some moral system, but whether or not it is actually good in accordance with one's self-preservation. The irrationality of the decision, however, does not follow as a result of emotional arousal or information; the irrationality stems, in accordance with Elster's notion, from the internal causal process which produces the experience – cognitive irrationality.

Hedonic adaption of preferences states that the social agent over time gets *habituated* to a certain hedonic experience, hence the new and pleasurable experiences lose their novelty, just as painful experiences lose their grip on the mind over time, because the agent starts to focus on other matters (Kahneman, 1999). This does not imply, however, that the pleasurable and painful ceases to be pleasurable and painful, what it implies is solely that they loses their novelty. In other words, hedonic adaption does not, necessarily, entail that social agents comes to want what they got. Now, if one takes into account aspirations, it becomes possible to extent the concept of hedonic adaption to *hedonic adaptive preferences* and *counterhedonic adaptive preferences* as discussed in chapter III. Hedonic adaption of preferences implies that the agent comes to want what she got, as it was described above with regard to dissonance reduction, or she comes to want what she got by devaluing what she cannot get, i.e. the sour grapes effect (Elster, 1983: 110; 2010: 221).

While beliefs based on preferences are irrational, preferences based on beliefs can be considered to be non-rational, insofar as preferences are not subdued to the same rational criteria as beliefs (Elster, 1983: 20; 2010: 221, see also the discussion on preference formation in the chapter III). Max Scheler (1915/2004) captures this psychological mechanism in his work “Das Ressentiment im Aufbau der Moralen” based on Friedrich Nietzsche's (1887/2000) concept from his work “On the Genealogy of Morals”:

Nietzsche (1887/2000) writes:

“The slave revolt in morality begins when *ressentiment* itself becomes creative and gives birth to values; the *ressentiment* of natures that are denied the true reaction, that of deeds, and compensate themselves with an imaginary revenge.” (Nietzsche, 1887/2000: 1.10, 472)

He goes on in a later section:

“It seems to me one is lying; a saccharine sweetness clings to every sound. Weakness is being lied into something *meritorious*, no doubt of it (...) and impotence which does not requite into “goodness of heart”; anxious lowliness into “humility”; subjection to those one hates into “obedience”...” (Nietzsche, 1887/2000: 1.14, 483)

Scheler (1915/2004) proposes a similar argument:

“Doch müssen wir ein anderes psychisches Gesetz dazu zu Hilfe nehmen. Bei Gelegenheit aller Art von starken Strebungen auf die Realisierung eines Wertes hin, bei gleichzeitig gefühlter Ohnmacht, dieses Streben wollend zu verwirklichen, z.B. ein Gut zu gewinnen, stellt sich eine Tendenz des Bewusstseins ein, den unbefriedigenden Zustand der Spannung zwischen Streben und Nichtkönnen dadurch zu überwinden, dass der positive Wert des betreffenden Gutes herabgesetzt, *geleugnet* wird, ja unter Umständen ein zu diesem Gut irgendwie Gegenteiliges als positiv wertvoll angestehen.” (Scheler, 1915/2004: 29)<sup>35</sup>

This understanding of the psychological mechanism of resentment leads Scheler (1915/2004: 4, 8) to describe it as *Selbstvergiftung* – self-poisoning – of the mind. Resentment is, perhaps, best described as a compound of two emotions, indignation and inferiority, in the sense that it is a form of *self-repression* of naturally occurring emotions such as envy. It is, in a sense, how the mind seeks refuge within itself by getting to terms with the reality facing it through reevaluation. The employee, for example, who do not obtain a reward she desired may be able to convince herself that she did not want it in the first place, or she may argue to herself and others that the other person who obtained the reward did something which violates some moral or social norm to obtain the reward – this psychological mechanism is what Elster (1999: 332; 2010: 223-224) calls *transmutation* understood as the transformation which occurs when “I did not obtain x” is transformed into “I did not want to obtain x”.

---

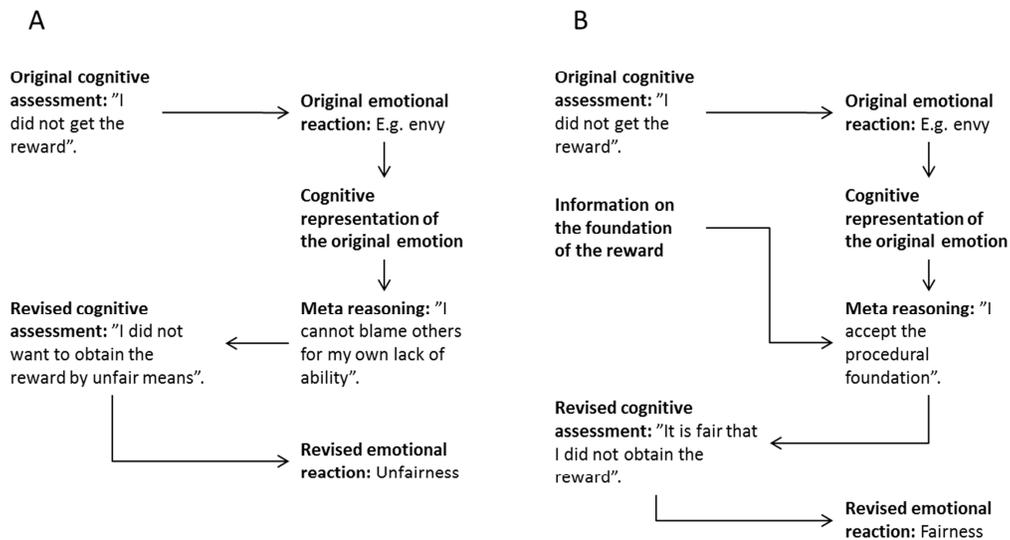
<sup>35</sup> The original quotation is applied because it is more elaborate on the psychological mechanism Scheler (1915/2004) perceived than the English translation: “However, we must introduce an additional psychological law. We have a tendency to overcome any strong tension between desire and impotence by depreciating or denying the positive value of the desired object. At times, indeed, we go so far as to extol another object which is somehow opposed to the first.” (Scheler, 1915/1994: 23).

Counterhedonic preference adaption, on the other hand, is of another kind, because whereas the former could be described as a case of the sour grapes effect, the present is best represented as a matter of “the grass is always greener” and, as such, it serves no actual cognitive purpose, besides perhaps making social agents’ miserable (Elster, 1983: 111). Furthermore, these devaluations associated with the sour grapes effect may also be counterhedonic when they turn into socially undermining processes (Elster, 2010: 224). Consider again the employee who starts to spread rumors about his colleague who obtained some reward; he may actually himself end up getting punished by his colleagues for engaging in such a behavior, if his accusations or statements regarding the other’s character is not based in any form of actual evidence visible to other parties than himself – his actions are, in a manner of speaking, self-defeating. The problem here is that “I did not obtain x” turns into a self-reflexive “I cannot obtain x”, while the intensity of the former is determined by the level of expectation associated with it, as also argued by expectancy theory, the latter hinges on the agent’s perceived abilities and, as such, it represents a threat to his will, from an external causal perspective, just as it invokes emotions associated with pride, and as argued above, when the question regarding distributional fairness turns into a matter of pride it invokes a different norm set. That is, when this self-reflexive transformation occurs, matters of distributional fairness turns into matters of perceived potentiality, and the emotional stress accompanying this sense of a perceived lack of potential is far greater than emotional stress accompanying the sense of unfairness, insofar as with regard to the latter the agent has no one else to blame than herself. To reduce this sense of psychological stress, the agent might view positive matters as being within her power, whereas negative matters are caused by external circumstances (Elster, 1999: 85). The behavior of the employees’ described in the Hawthorne study who punished one another for being too productive and made matters regarding effort into matters of solidarity, or the employees described in Lysgaard’s study who turned this into a matter of workman’s pride, may then be explained along these lines, because feelings of inferiority easily turns into these misrepresentations of self-reflexive emotions as matters of solidarity or unfairness. That is, one cannot blame others for one’s own lack of ability or others for being better, but one can blame them for violating the social norms or the system for being unfair. This might, of course, be described as matters concerning the interaction between esteem and self-

esteem which are misrepresented as claims on unfair behavior by either the employee who received the reward by violating the social norms or the organization (Elster, 1999: 94, 95).

This also implies that the judgment produced is, at least in the perspective developed here, irrationally produced in the sense that it is not based on *real* evidence of unfairness. A theory centered upon assigning blame, then, is problematic if it does not take into account the nature of these judgments. However, as argued by Folger and Martin (1986) social agents have a tendency to accept distributions of rewards if they are informed about its foundation, implying that perceived procedural fairness in some way offsets this sense of inferiority, which might imply that in cases where the foundation of a reward differentiation is properly explained and accepted, the agent's possibility of formulating a plausible mental story in which she did not get the reward due to some unfairness is less likely. The difference between the two ways of reasoning may be illustrated as follows:

Figure 30: Mental processing of fairness issues (Inspired by Elster, 1999: 99).



In the figure above, the mental processing of fairness issues is sketched to explain the difference between hedonic adaption and acceptance of reasons. The process described on the left hand side (A), shows how the event which triggers the process is firstly framed as an event in which the agent did not get the reward she expected, this belief triggers the emotion, and the emotion combined with the belief triggers a cognitive rep-

resentation of the emotion within the agents consciousness – for example “I am envious” or “I feel inferior” – in a sense, it is the agent’s confrontation with her own emotion. This confrontation within the consciousness of the social agent, then, triggers a meta-reasoning based on the questions “why do I feel the way I do?” and “what can I do to change how I feel?” leading to a revised cognitive framing which eventually leads to a revised emotional reaction. The difference between the conclusion of the meta-reasoning process and the revised cognitive assessment illustrates the irrationality implied in preference reversals. The cognitive process illustrated on the right hand side (B) is similar, although the introduction of externally produced reasons influences the process of meta-reasoning and, as such, aligns the conclusions of meta-reasoning with the revised cognitive assessment. Acceptance, as such, is centered upon the acceptance of externally produced reasons, whereas hedonic adaption is centered upon purely internally produced reasons. Acceptance may further be split up into two subcategories, acceptance with voice based on autonomy (as discussed above), and acceptance without voice which is similar to the process of hedonic adaption, because in this case the employee can either accept the consequences or leave the organization. Acceptance without voice, in other words, is similar to what Merton (1968) described as the anomic consequences of relative deprivation (see chapter VI).

#### **7.4 On the Foundations of Expectations**

The past and the normal state, then, it seem generates the cognitive point of reference for one’s present experience, insofar as the present experience it brought about by comparison. Deviations from this point of reference, be they positive or negative, are amplified because status quo takes prerogative over the present and, as such, the normal state and the past takes on values which are overvalued – that is, the evaluation of the past and the normal state is asymmetric to the evaluation of the present or future. This overvaluation of the past is also present in what is often referred to as the *endowment effect* which describes how something which is already in the agent’s possession (endowment) takes on a higher value (Kahneman, 2011: 290). The endowment effect emerges as a consequence of a comparison between “what I had” (past) and “what I have” (present) or perhaps “what I have” (present) and “what I stand to gain/lose” (future), and due to

loss aversion disadvantageous comparisons looms larger than advantageous comparisons, implying that the pain of giving up overshadows the pleasure of getting (Kahneman, 2011: 293). As a consequence of this, sunk costs matters and opportunity costs are undervalued. The endowment may, then, be viewed as representing a particular hedonic value to which one is habituated and which as a consequence comes to stand as a norm, although this should not be understood as representing a hedonic zero. The endowment effect represents, in other words, how the *seller's* subjective price ascribed to status quo may be higher than the *buyer's* price – meaning that if the seller is to part from what is now within her possession, she expects a price which matches her subjective price – that is, if she is to move from one status quo to another, she must be compensated in a manner which enables this change to occur without her experiencing it as a loss. To demonstrate the endowment effect within employment relationships consider a standard Nash bargaining game:

This game have three bargaining problems (1) Bargaining over the wage ( $t$ ); (2) Bargaining over the job content. The employer wants the employee to take on a job with many different tasks ( $x = 1$ ), while the employee wants a job like his current job with few and specialized tasks ( $x = 0$ ); and (3) the parties bargaining over whether they should initiate an employment relationship or not?

The employer's utility from the job with many different tasks is 50,000, while the employee's is 12,000. The employer's utility from the specialized job is 45,000, while the employee's is 15,000. The parties' status quo utility is:  $d_{employee} = 20,000$  and  $d_{employer} = 15,000$ . This yields the following bargaining set with associated payoff vectors:<sup>36</sup>

*Initiate an employment relationship with a specialized job ( $x = 0$ ):*

$$\text{Payoff Employee } (x = 0) = 15,000 - (15,000 - 12,000)x + t$$

$$\text{Payoff Employer } (x = 0) = 45,000 - (45,000 - 50,000)x - t$$

$$V(x = 0) = (15,000 - 3000x) + (45,000 + 5,000x) \rightarrow 15,000 + 45,000 + 5,000x - 3,000x \rightarrow 65,000 + 2,000x = 60,000$$

---

<sup>36</sup> For simplicity I have chosen to transform the convex curves applied in the normal Nash solution into simple linear curves.

$$\text{Surplus} = 60,000 - (20,000 + 15,000) = 25,000$$

*Initiate an employment relationship with a non-specialized job (x = 1):*

$$\text{Payoff Employee (x = 1)} = 15,000 - 3,000x + t$$

$$\text{Payoff Employer (x = 1)} = 45,000 + 5,000x - t$$

$$V(x = 1) = (10,000 - 3,000x + t) + (45,000 + 5,000x - t) = 60,000 + 2,000x = 62,000$$

$$\text{Surplus} = 62,000 - 35,000 = 27,000$$

Now, the efficient solution is for the parties to initiate the employment relationship and to agree on the non-specialized job, because this yields the highest overall surplus, although the employee's payoff decreases this is offset by the increase in the employer's payoff and the employee can be compensated through  $t$ . The parties in game, however, are not actually bargaining over efficient surpluses, rather they are bargaining over how the surplus should be split insofar as each player is trying to maximize his/her payoff and by not initiating the employment relationship each player can get their respective status quo utility, thus no *rational* player would agree to  $U < d$ . The proportional division of the surplus is expressed as bargaining power and mathematically expressed in the bargain weights, hence if  $\pi_{\text{Employer}} = 2/3$  and  $\pi_{\text{Employee}} = 1/3$ , then:

$$u_{\text{employee}} = 20,000 + \left(\frac{1}{3}\right)(27,000) = 29,000$$

$$u_{\text{employer}} = 15,000 + \left(\frac{2}{3}\right)(27,000) = 33,000$$

The  $t$  which enables the required split of the surplus to occur can now be calculated:

$$29,000 = 15,000 - 3,000(1) + t \rightarrow -t = 12,000 - 29,000 \rightarrow t = 17,000$$

The conclusion, then, is that the employee should agree to the non-specialized job and be paid 17,000. As this simple bargaining game illustrates, the status quo utility determines the surplus and also has consequences for the compensation, just consider what happens if the employee, for example, really likes his current job,  $d_{\text{employee}} = 40,000$ , the total surplus would decrease by 20,000, and the employee would be better off just

staying in his current job, because  $U < d = 22,300 < 40,000$ . How much should the employee's pay increase to make him consider taking the job? First of all, his proportion of the surplus would have to increase:

$$20,000 + \pi_{employee}(7,000) > 40,000 \rightarrow 7,000\pi_{employee} > 40,000 - 20,000 \rightarrow$$

$$\pi_{employee} > \left(\frac{20,000}{7,000}\right), \text{ hence } t \text{ should be set to: } 40,000 > 15,000 - 3,000 + t \rightarrow -t >$$

$$15,000 - 3,000 - 40,000 \rightarrow t > 28,000.$$

This implies that if the status quo utility increases the employee's proportion of the surplus also increases and the pay needed to compensate the employee for making the change increases. The pay level, in other words, illustrates the price which the employee considers as fair should he be willing to make the change required, in the sense that it compensates him for his loss of utility associated with deviating from status quo.

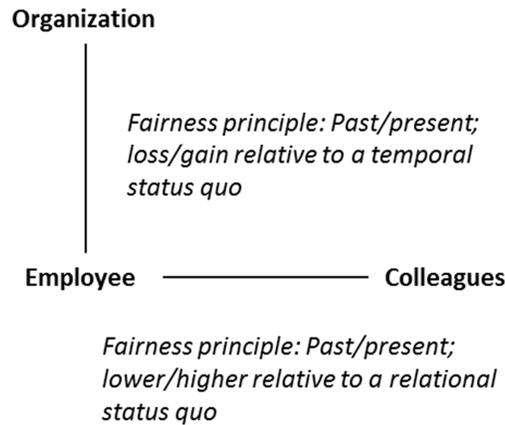
Consider now a corollary of this. From what was argued above, the parties in an exchange were entitled to their status quo level. Does this also apply in a bargaining game? Simon Gächter and Arno Riedl (2005) studied this by conducting a series of experiments with bargaining over the size of the bargaining weights. Initially, the fictive organization described in their experiment, paid 1,660 to high-performers, and 830 to low-performers - equivalent to  $([1,660/2,490]*100) \frac{2}{3}$  to high-performers and  $([830/2,490]*100) \frac{1}{3}$  to low-performers (Gächter and Riedl, 2005: 251). Due to economic recession the organization's pay budget was reduced to 2,050 making the previous lump sums infeasible solutions to the bargaining game. The participants in the experiment were now to bargain over how to split the lower salary budget between them. However, before initiating this game, the participants filled out a knowledge-test with some simple questions and they were informed about how good they did in terms of ranks – this was done to preserve the distinction between good and low performers (Gächter and Riedl, 2005: 251). Additionally, the participants were asked what they considered, from a neutral stand point, to be a fair distribution between high- and low-performers in an organization experiencing economic recession (Gächter and Riedl, 2005: 252). That is, if an organization is experiencing economic recession is it, then, still obligated to pay the high-performers their status quo proportion,  $\frac{2}{3}$ , or does this

decrease of the budget push the fair split between low- and high-performers closer to 50/50?

What the authors found was that only about 0.10 percent of the low-performers preferred an equal split, while the majority of the low- and high-performers preferred a 65-67.5 percent split to the winners – implying that the equity rule was found to be preferred by both parties. The average fair split considered by the high-performers was 64.0 percent and 61.6 percent by low-performers (Gächter and Riedl, 2005: 253). Furthermore, the study also showed that when making opening offers, the high-performers tended to open with an offer above what they found to be fair from a neutral stand point, whereas the low-performers tended to open with an offer lower than what they from a neutral point judged to be fair (Gächter and Riedl, 2005: 254). The agreements reached by the parties, however, were positively correlated with the fairness judgments, while disagreement was only indirectly related to fairness judgments (Gächter and Riedl, 2005: 257).

What this experiment seems to illustrate is that entitlements matters in bargaining games, because the high-performers ended with a percentage share of the new budget close to the percentage share they had under the previous salary budget. Just as Messick and Sentis (1979) argued, in regard to individual choice, social agents tend to choose what they consider to be fair as opposed to what they would prefer. However, even though the high-performers' percentage share did not change much, they would with the lower budget receive a lower amount, implying that whereas they might be entitled to a particular percentage, they were not entitled to a particular amount. The lower percentage share obtained can, of course, in the relation between the employee and the organization be viewed as a minor loss, just as it from a horizontal perspective (high-performers vs. low-performers) can be viewed as a slight change of relational structure. Nevertheless, combining these two dimensions one can argue that entitlements exists as claims related with temporal phases, past vs. present, and that these relates to the vertical dimension of the relationship between the organization and the employee; and as claims related to the horizontal relational structuring of the organization. This can be illustrated as follows:

Figure 31: The two dimensions of fairness.



Kahneman, Knetsch, and Thaler (1986a; 1986b) and Gächter and Reidl's (2005) draw attention to an important matter, morally founded expectations. While the first group of authors discusses moral judgment as a matter of framing and, as such, discusses the cognitive foundation of moral judgment; the second group of authors discusses the moral normative foundation. Furthermore, the first group of authors call these morally founded rights, entitlements, while the second group calls them claim rights. Perceived entitlement is related with a temporal status quo, while claims are related with recognized claim rights. As such, the cognitive foundation is related with the activation of beliefs which activates certain emotions related with moral judgments – that is, framing occurs as a result of the interaction between the beliefs held by the social agent and the context which consists of a limited space of possible explanations. This context activates beliefs within the social agent which then activates emotions. In that sense, the moral judgment is partly cognitive grounded in the beliefs activated, and partly grounded in the emotions which it invokes within the individual. This is, of course, different from the moral intuitionists who, based on Hume's (1739/1985: 509, 510) moral philosophy in which he claims that emotions – the experience of pleasure or disgust – to be the foundation of morality, argues that the foundation of moral judgment is the emotional experience stirred up within the individual as a reaction on something – as a kind of automatic cognition (Haidt, 2012: 46, 47). While this is part of a larger discussion between, on the one hand, moral philosophy and moral psychology, and, on the other hand, between the rationalist tradition and the intuitionist tradition in moral philosophy; it might just be argued that emotions are invoked within the individual as a reaction on

how something deviates from an already held expectation which might be either *morally founded* as a principle (see my discussion in chapter III), socially founded in expectations on how other people act and react, or cognitively founded in expectations on what is considered to be the *normal* course of events. What these three bases of expectations have in common is that deviations stir up emotions and that the experience of this deviation is related with the *direction* (positive/negative) and *magnitude* of the deviation from some experience; and the *certainty* (cognitively founded in the distinction between normal/abnormal) or *legitimacy* (socially and morally founded in the perceived recognition of the right – the agents’ beliefs on what is socially and morally within their rights, and supported by some kind of social institution, that is, the agents believe that other social agents also recognizes their right). However, whereas cognitive normative expectations are primarily related with experience and memory, moral and social expectations are accompanied by a claim of legitimacy and collective recognition – that is, they make a claim on other social agents. Of course, one could also, as Folger and Cropanzano (1998; 2001) does, argue that deviations are caused by other social agents’ deviations and, as such, the demarcation between the different kinds of expectations may often be blurred in real life situations. Nevertheless, theoretically, at least, it should be expected that deviations from cognitive normative expectations to be associated with confirmation/disappointment (Elster, 2007), whereas expectations which make a claim on other social agents’ actions and beliefs might be accompanied with a different set of emotional experience – not so different from the framework proposed by Boudon’s (2001) reformulation of Smith’s theory on the impartial spectator. That is, one should expect a qualitative difference between being disappointed when something deviates from what is expected, and not being recognized as being in possession of the moral or social rights to claim something.

## 7.5 Tentative Conclusions

The aim of this chapter has been to discuss the third dimension of the cognitive structuring of outcome claims, temporality and its psychophysical manifestation in comparison based on hedonic experience – recalled and presently experienced. In the previous chapters I have argued that the outcome claim is also founded on a comparison between ac-

tual and potential worlds and relational proportionality. Outcome claims, in other words, emerges from normative expectations on: (1) a *normal* world equal to the potential world which should have been realized had everybody acted in accordance with expectations; (2) a world in which relational proportionality is respected and reflected vertically in the exchange between input and outcome and horizontally in the maintaining of a proportional difference in the input/outcome ratio equal to differences in input; and (3) a world in which comparisons between past and present cannot be associated with a hedonic experienced loss.

A fair vertical exchange is one that maintains relational proportionality and qualitative equality between past and present or before and after, just as it is one which does not deviate from one's normative expectations on what characterizes a normal world – a fair world. The unfair exchange is one that deviates. Besides these outcome claims, fairness at a more general level is also related with a fair treatment – being treated fairly and behaving fairly towards other social agents. In vertical relations, the fair treatment manifests itself in the exchange and its symbolic transformations into signs on respect of the others rights. Vertically as well as horizontally, then, the fair treatment of others includes the acceptance of other-regarding obligations a symbolic act accompanied by the hope on one's own rights being respected – that other social agents accept their obligations, take it upon themselves to act in accordance with these, just as one must take it upon oneself to act in accordance with these. This deontance inspired model can be contrasted either with the Hobbes-Coleman rational model or Fehr's model on reciprocity, in the sense that, here, the agents' behavior are contingent upon the behavior of the others – either in an egoistic-rational manner or in an irrational-social manner.

In accordance with what has just been argued, the after should be equivalent to the before from which it, in a temporal-logical manner, emerged – thus, the primary principle of fairness is that one is not allowed to impose a loss on another individual. That is, entitlements emerge, as described above, from past interaction or exchange. Loss is, of course, a relative term which is contingent upon choice of reference point. Whether or not a loss has occurred is a matter of coding – that is, compared with status quo has a loss occurred – and the coding process (the process of sense making) is influenced upon the framing of decision context – the temporal structure of the course of events and the

opportunities faced by the decision-maker. How a matter is coded is, in other words, relative to how it is framed by the context – how it appears to the agent in the situation. Hence, depending on a particular decision context is framed; the agent may prefer A to B in one situation framed in a particular way and B to A in a situation framed in another way. Preferences are not just context dependent and related with the opportunities faced by the agent, they are also contextually constituted. In a similar way, judgment on issues related with fairness is vulnerable to framing effects. A judgment, then, is not based on careful processing of information and evaluations of values; rather, it is centered upon a particular manner of interacting with the context – how relevant information is retrieved from the context, encoded in a particular manner, and transformed into a particular judgment. Furthermore, to enable the easiest production of a judgment, different decision rules are invoked. A judgment on fairness emerges in this perspective, then, from an interaction between the relevant retrieved information, the coding process, and the decision rule. In that sense, a judgment on fairness is not based on reasons or principles – rather it is simply a matter of cognitive processing of information. The judgment on fairness may be viewed as being in opposition to economic rationality because it is centered on a different underlying system of values, just as it may be viewed as being in opposition to the rational tradition because the judgment is not based on a deliberate process of careful assessment of information. Is the judgment then economically irrational and irrationally produced? If a judgment is irrational it would imply that something had gone wrong in the process of production, insofar as rationality cannot be equated with the production of correct responses – all that is claimed is that a rational decision is based on strong reasons. A response which deviates from a particular value system cannot be viewed as a sign of evaluative irrationality, just as a wrong response cannot be viewed as a sign of irrationality. The individual may be wrong, but still have strong reasons to believe that she is right – thus, as long as her judgment is based on strong reasons she is rational – she is not rational because she has not been able to produce a correct response since this might be based on wrong reasons.

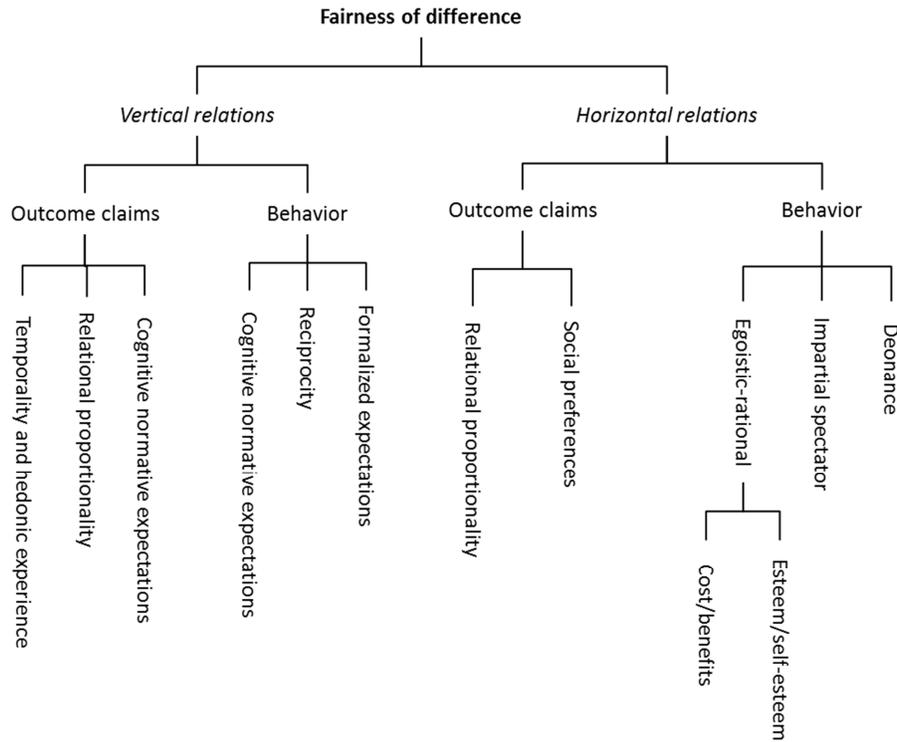
From a first-person perspective, hedonic experience as the basis of fairness seems problematic, because as argued by Plato that which is pleasurable may not be good and neither may that which is painful be wrong, implying that some pleasurable desires are accompanied by the painful experience of it being wrong. This led Plato and later Kant

to argue that the basis of judgment cannot, solely at least, be hedonic experience at least not without the aid of reason. Hume, in opposition to this, argued that the foundation of judgment was not reason, but attraction and repulsion equivalent to the hedonic experience of pain and pleasure. Whereas the former tradition may be criticized for its focus on the production of universal judgments, the latter tradition may be criticized for its focus on experience, insofar as this, as argued above, is vulnerable to framing effects. Besides the epistemological problem, one may also argue that hedonic experience is problematic to apply as the foundation of choice due to hedonic adaption – that is, human beings has the capacity to adapt their qualitative experience to the situation facing them. The painful experience, for example, does not cease to be painful, but it loses its novelty. Moreover, due to hedonic adaption prolonged exposure to painful experiences reduces the qualitative intensity of the experience, whereas as short term experiences are recalled as more painful. Hence, the prolonged experience may be preferred to the short term experience. If the least painful option is chosen in order to minimize the painful experience this choice may from a Hobbesian perspective be irrational, insofar as this choice is against the primary principle of rationality, just as it is in opposition to the principle of choice itself.

Besides this problem associated with hedonic experience, hedonic adaption may also be problematic to apply as the foundation of the experience of unfairness, insofar as the agent who is prolonged exposed to unfairness may adapt her level of aspiration and qualitative experience to the situation, just as the opposite may also occur if the agent maintains her perception that she could always do better. In the latter situation, an experience of unfairness is always present, whereas in the former situation, unfairness is seldom experienced because the agent has internalized the experience and adapted her personal sense of worth to the situation. In that sense, to apply hedonic experience as the foundation of fairness is problematic not, however, due to the individual's ability to collect information, but simply due to cognitive irrationality – the very production of experience.

From what has been discussed so far, the following hierarchical taxonomy seems to emerge on fairness of difference:

Figure 32: Hierarchical taxonomy of fairness of difference.



From a vertical perspective, the fair treatment can be divided into a section on behavior and a section on the constitution of the outcome claim. In a similar manner the horizontal dimension can also be divided into a section of behavior and one on the constitution of the outcome claims. Fairness cannot be limited to outcome claims, insofar as the interaction cannot be limited to exchange relations, but encompasses a social dimension related with how the agents act towards one another. Vertically, a claim is constituted on by the three bases discussed in beginning of this section. Horizontally, an outcome difference is justified based on social preferences and relational proportionality. Fair behavior, in horizontal relations, is related with the acceptance of the claim rights of other social agents – the self-commitment to respect and take it upon oneself to respect these and hope that one’s fellow human beings also will accept their obligations. In juxtaposition to this, the agent may also accept her obligations due to a cost/benefit analysis or due to concerns about esteem/self-esteem. Regardless of the motive, what is captured in this category is the concept of social rationality – the behavior which enables the contiuance of social relations. Social rational principles can be identified in both relations.

## Chapter VIII

### Self-Differentiation and the Need to Belong

#### 8.1 Introduction

In the previous chapters I have been discussing the constitution of outcome claims as well as the relational ties which bind agents to one another. The aim, in that sense, has been to understand how social rationality enabled the continuation of social relations – that is, how behavior in accordance with governing social rules enabled the existence of social relations. Moreover, I also argued that the socially optimal decision is one that maintains the existence of these social relations. Social rationality cannot, however, be limited to rules of interaction, insofar as claim rights related with outcome and outcome differences are also socially produced. From a vertical perspective, the foundation of claim rights has been identified as being contingent upon: temporality, cognitive normative expectations, and proportionality. Horizontally, these claims rights have been associated with maintaining relational status quo and social preferences. Vertically, then, the claim right represents the foundation of fair expectations, whereas horizontally, claim rights are related with differences and why these are accepted as fair. The focus has, as such, been on understanding the constitution of fair differences and what binds the agents to one another.

In the present chapter, I will shift from a focus on difference to a focus on differentiation in order to analyze the social dynamics of differentiation. In chapter VI, I argued that from a horizontal perspective, social agents cannot blame one another for bringing about particular outcome states, insofar as no real causal relation can be identified. Moreover, in this chapter, I also argued that whereas social agents cannot be blamed for bringing about certain compensatory difference, they may be blamed for producing certain actions which indirectly causes these differences. The outcome difference may, in other words, follow from an act of differentiation. In chapter II, I identified how social norms may be introduced to govern these acts of differentiation. In the present chapter, I will argue that one should distinguish between quantitative differentiations related with, for example, the supply of a higher level of effort; and qualitative differentiation related with differentiation through non-quantifiable choices, see also

chapter II. Differentiation, I will argue in the present chapter, is opposite to the primary principle of social rationality – to maintain social relations – in the sense that the aim of differentiation is to produce differences – to differentiate one from those around oneself. This production of self-differentiation may, however, be problematic from a social rational perspective, insofar as it seems not only be in opposition to the primary principle, but also in opposition to the basic social need – the need to belong. In a simplified manner, then, one may distinguish between social rationality which binds and pull together and acts of differentiation which has as their aim to push apart and estrange the agents from one another. To understand this, one needs to understand how social relations are tied together by the existence of and interaction with social norms, because differentiation may be in opposition to social norms, but not necessarily in opposition to moral norms – thus, the focus here is on the third constitutive part of social rationality, social norms.

## 8.2 Social Norms

To pursue this line of reasoning, I turn to Elster's (1989b; 2007; 2009a) theory on social norms. According to Elster (1989b: 99-100; 2009a: 196) a social norm is social, because it is shared – and known to be so – just as it is socially reproduced by the social agents taking it into consideration when action, and following from this, it is partly maintained by material social sanctions and partly by non-material social sanctions. As such, a social norm is accompanied by the agent's social expectation that other social agents expect her to conform to this norm (act in a given way), and react to a particular behavior (norm violation) in a particular way which will be harmful to the agent (Elster, 2009a: 196). In that sense, norm following is conditional on the agent's possibility of being observed by other social agents – implying that a social norm is solely activated in the presence of other social agents (Elster, 2007: 104).

According to this perspective, then, social agents does not conform to social norms, because they view the action as right, in the sense they believe that they and others *ought* to conform to this action, thus it is not accompanied by what Kant calls a tacit claim on the world. Such a claim, of course, would change the nature of the norm from being a social norm to a moral norm (Elster, 1989b: 101; 2007: 104-105; 2009a: 196).

Additionally, the decision to follow a norm is non-consequentialistic, insofar as the agent does not adhere to the norm due to a belief about it having favorable consequences (Elster, 1989b: 101). It is however accompanied by the agent's quasi-normative expectation on others expecting her to conform and that they are willing to sanction her if she does not. Although quasi-normative it is derived from social expectations, therefore a perhaps more appropriate definition would be social normative expectations.

In addition to not being moral norms, social norms are not legal norms, because these are based on objective laws and sanctioned through a formal process, just as the motive behind following these often are purely consequentialistic (Elster, 1989b: 101; 2007: 357; 2009a: 197). Neither are social norms private norms nor habits, while the former refers to self-imposed norms, the latter refers to, often unconsciously, frequently repeated actions (Elster, 1989: 102, 104). Social norms should also be distinguished from tradition understood as the *mindless* repetition of previous generations' actions, just as social norms should also be distinguished from equilibria emerging in evolutionary games (e.g. focal points or optimal long-term strategies such as TIT-for-TAT in the prisoners' dilemma), and social conventions – for example, always wear a black suit when attending business meetings (Elster, 1989b: 104; 2007: 357; 2009a: 198). Social norms also differs from what Kahneman and Miller (1986) called norms, because social norms are not cognitive expectations, but purely social in nature. Finally, social norms also differs from quasi-moral norms, because whereas norm following under a social norm is conditional on being observed by another social agent, norm following under a quasi-moral norm is conditional on the agent herself observing the actions of another agent – reciprocity, for example, is a quasi-moral norm, insofar as the agent's choice on whether or not to conform is contingent on whether or not the other party conforms to the norm (Elster, 2007: 104; 2009a: 196). By adding the last point of difference, Elster (1989b; 2007; 2009a) is distinguishing his theory from other currently dominant theories on social norms within social science today, for example, Fehr and Gächter (2000) and Bicchieri's (2006) respective theories.

According to Fehr and Gächter (2000: 166) a social norm is an observed behavioral regularity, based on a collectively shared belief on how one ought to behave, and

this triggers the enforcement of the prescribed by informal social sanctions. Fehr and Gächter (2000) argue further:

“...a social norm can be thought of as a sort of behavioral public good, in which everybody should make a positive contribution – that is, follow the social norm – and also where individuals must be willing to enforce the social norm with informal social sanctions, even at some immediate cost to themselves.” (Fehr and Gächter, 2000: 166)

According to this perspective, then, norm following is centered upon a social expectation regarding how the agent herself and others are supposed to act, and the second-order belief that others also expect the agent to behave in a given manner. Moreover, these social expectations are supported by the normative expectation on how one ought to act, because it is within the collective interest to secure that this social norm is observed. Fehr and Gächter (2000) it seems, then, pursue a perspective similar to Coleman’s (1990), although without the instrumental dimension; rather than appealing to the consequentialistic dimension of norm violation, they focus on the obligation which is attached to the norm.

Following a similar line of reasoning, Bicchieri (2006: 8) argues that a social norm is a public and shared norm supported, if at all, by informal sanctions. Violation of the norm may invoke shame and remorse in the transgressor and this may enforce the agent’s desire to conform, albeit it can never be the final explanation. Social norms are centered upon conditional social and normative expectations about how the agent herself and others are supposed act, and second-order beliefs on how others expect the agent to behave. According to this perspective, then: (1) the social agent knows that the norm exists and she conforms to it in certain situations in which she judges it to be appropriate; (2) the agent prefers to conform to the norm in these situations if (a) she believes that a satisfactory part of the other social agents also conforms; and (b) the agent believes that a satisfactory part of the other social agents expects her to conform to the norm in the particular situation or she believes that this satisfactory part of the social agents who expect her to conform also are ready to sanction her if she does not conform (Bicchieri, 2006:11, 12, 13, 15). Bicchieri’s (2006) perspective, it seems, is closer related to Elster’s (1989; 2007; 2009a) perspective than Fehr and Gächter’s (2000) is, with the exception, however, that according to her perspective social norms are centered upon a double set of conditional expectations: (1) the agents conformity is conditional

on other agents conforming to the norm; and (2) the agent's expectation about other social agents expecting her to conform to the norm. In Elster's (1989b; 2007; 2009a) perspective, norm following is solely conditional on the presence of others and the agent's expectation on that they are willing to sanction her and, as such, the agent's behavior is unconditional on whether or not the others themselves are conforming to the norm. Although, one can argue that if nobody seems to be obeying the norm, why the agent then should do it, it seems plausible that even though these other social agents do not appear to obey the norm, they might still hold it, thus they might still be willing to sanction the agent. Furthermore, it also seems that Bicchieri's (2006) perspective turns every social interaction into a prisoners' dilemma coordination game in which there is only one optimal strategy, to conform until others deviate. In other words, it seems as she presupposes that only one social norm exists in a particular situation – an understanding which is also supported by her use of psychological scripts:

“To say that preferences (and beliefs) are context-dependent is to say that they are sensitive to situational cues and the subject's interpretation of these cues. It follows that particular preferences and beliefs may not always be activated; rather, they are the result of an interpretation of specific cues, a categorization of the situation based on those cues, and the consequent activation of appropriate scripts.” (Bicchieri, 2006: 57)

She goes on later in the text:

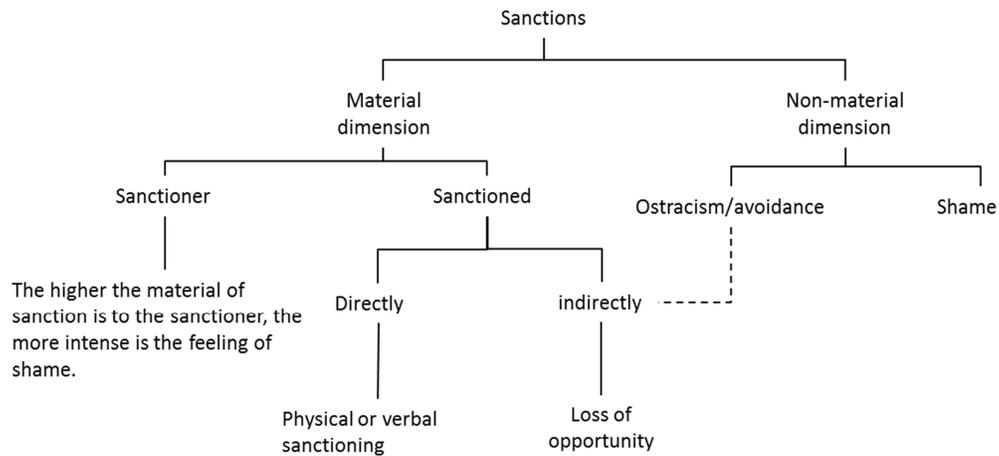
“When people use scripts, they know what to expect of each other. They need not be acquainted with each other or know of each other's past performance. Their expectations are grounded in the certainty that, *if* indeed script *s* is being enacted, *then* actions  $a_1, \dots, a_n$  will follow.” (Bicchieri, 2006: 96).

In that sense, the existence of social norms reduces the ambiguity of the situation by only allowing a particular norm and a particular subset of actions to be permissible. This also entails that in situations framed as a particular kind, the agent will always act in a similar manner. Although Elster (2007: 185) proposes a similar perspective when he argues that actions are better explained with reference to the situation in which they occur than with reference to personality, he also adds that situations are not unique, but gives rise to many different explanations – an explanation which is close to Merton's (1957: 110-111) explanation of the situation as one which invokes different roles with different sets of actions from which the social agent can choose. If only one norm existed, then, conflicts over norms would not occur, insofar as it would solely be a matter of

continuing the coordination game until a shared norm was identified, although this might be often be the case, it seems, at least, sensible to imagine that several different norms exists and applies to various situations – an understanding of social norms which also seems to move the existence of social norms from situations and the framing of such to actions and the values which they represent, and which comes to be espoused when these are carried out by the agent (Elster, 2007: 353). In that sense, social norms are associated with particular actions not situations, however, some actions are more frequently performed in certain situations, thus one cannot, entirely at least, remove the situation from consideration. In some way, then, Elster's (1989b; 2007; 2009a) perspective seems closely related with Merton's (1968) concept of social norms presented above. Bicchieri (2006), nevertheless, touches upon a crucial point, when she emphasizes that the agent must believe that the norm applies, because can a social agent be sanctioned for violating a social norm she did know existed? Elster (2009a: 200) seems to argue that this is one of the ways in which social agents become aware of norms – that is, initially, the agent may not be aware of the fact that she has violated a norm, but she learns that this is so when she continues to do it and observes other social agent's reactions on her behavior. This follows, it seems, as a logical consequence of his idea about norms being activated by the presence of others, insofar as actions comes to represent certain values only, or at least primarily, when performed in front of others – an idea which seems to be closely related to Boudon's (2001) theory about the impartial spectator.

Before moving on to the contrast between Elster's (1989; 2007; 2009a) and Fehr and Gächter's (2000) perspective, it seems appropriate to present the last part of Elster's (1989b; 2007; 2009a) perspective. Social norms, he argues, gain causal efficacy because they are associated with sanctions which may either be material or non-material, see the figure below:

Figure 33: Illustration of the analytical logic of sanctions.



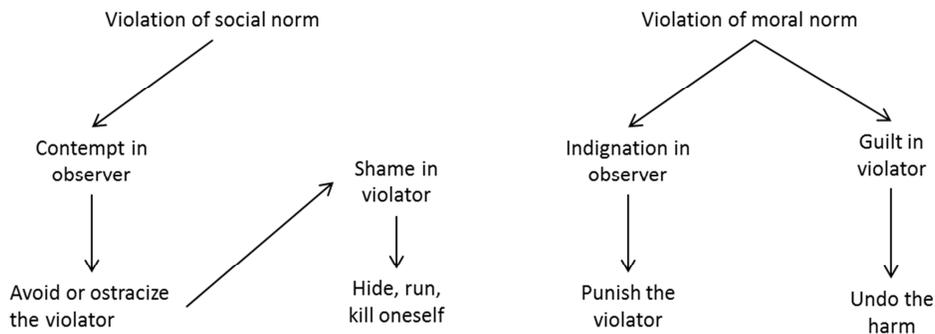
As the figure illustrates, a sanction may either be material or non-material, and if material it may either be directly associated with a material loss or indirectly through loss of future opportunity. The material dimension may also be related with the material costs incurred by the sanctioned agent, described above, or the material costs incurred by the sanctioning agent – a matter I return to below. If, on the other hand, the sanction is non-material in kind, then, it might produce shame in the violator or the sanctioning agent might avoid interaction with the agent which might lead to an indirect material loss of opportunity (Elster, 2007: 354). The non-material dimension, however, should not be conceived as a matter of internalized sanctioning, at least, not in the sense that the agent will feel shame for having violated a norm if her transgression was not observed, although Elster (1989b) in his early account seems to allow this:

“...norms are sustained by the feelings of embarrassment, anxiety, guilt and shame that a person suffers at the prospect of violating them, or at least at the prospect of being caught violating them. Social norms have a *grip on the mind* that is due to the strong emotions their violation can trigger. I believe that the emotive aspect of norms is a more fundamental feature than the more frequently cited cognitive aspect. If norms can coordinate expectations, it is only because the violation of norms is known to trigger strong negative emotions, in the violator and in other people.” (Elster, 1989b: 100, italics in original)

What Elster (1989b) seems to arguing, here, is that norms gain their causal impact on actions through the emotions they trigger within the agent – what he refers to as the grip that norms have on the mind of the agent. This seems to be in opposition to Bicchieri’s

(2006) perspective, because he argues against the cognitive dimension as being the most important. Of course, Bicchieri (2006) does not solely consider the cognitive dimension; rather, she takes into consideration the cognitive, social, and normative dimension of norms. Nevertheless, in her book she criticizes Elster’s (1989b) perspective for being too focused on the internalized dimension of social norms, in the sense that if the norm had already been internalized the social situation would not matter, it would just, as she argues, be “...*preaching to the already converted...*” (Bicchieri, 2006: 75). Be that as it may, Elster (2007; 2009a) in his more recent restatements of his theory, seems to be focusing more on the norm being invoked by the presence of others – which, of course, seems to, at least to some degree, suggesting that if the norm exists as a particular value which the action takes on when carried out in the presence of others, then, it presupposes some kind of internalization to have occurred at some previous time. The theory proposed by Elster (2007; 2009a) can be illustrated in the following manner:

**Figure 34: Violations of social norms and emotions (Elster, 2009a: 197).**



The figure illustrates the analytical logic behind Elster’s (2009a) reformulated theory and the difference between moral and social norms. In some way, this reformulation solves the seemingly theoretical inconsistency between his originally proposed theory on norms and his current theory, because as he argues in his work on emotions, to add only the internal dimension of emotions would solely be to argue that the material model proposed by, for example, Coleman (1990) was incomplete (Elster, 1999: 146). He continues this argument by stating that the material perspective is not incomplete, but wrong:

“I believe it to be false, because sanctions do not work by imposing material losses on their targets. When I refuse to deal with a person who has violated

a social norm, he may suffer a financial loss. Far more importantly, however, he will see the sanction as a vehicle for the emotions of contempt or disgust and suffer shame as a result.” (Elster, 1999: 146)

What Elster (1999) seems to be arguing here, is that the sanctioning of the norm violator does not owe its impact on the norm transgressor to the material loss she incurs; rather it owes its causal impact to the emotions which the action itself comes to symbolize, thus a sanction can be viewed as a vehicle for communicating this emotion (2007: 355). Additionally, it functions as a vehicle of communication because it is often spontaneous – like when A wrinkles her eyebrows in disapproval of B’s action – and as such it should be contrasted with deliberate sanctioning or sanctioning based on selfish-motives, for example, the fear of being sanctioned herself for not sanctioning the norm transgressor insofar as these would, according to Elster (2009a: 201) at least, produce adverse consequences. Moreover, if those who sanctioned did so only to avoid being sanctioned themselves, it would entail infinite regress, insofar as someone should be willing to sanction the non-sanctioning agent, second-order sanctions, just as somebody should be willing to sanction the non-sanctioning non-sanctioning non-sanctioning agent, third order sanctions, and so on (Elster, 1999: 155; see also the discussion of Coleman’s norm theory in chapter III). Hence, some social agents, at least, must do so unmoved by a secondary motive.

Elster (1999: 139) refers to contempt and shame as social emotions, that is, emotions which are only triggered by beliefs that make a reference to other people. Furthermore, these emotions are interaction based, and as such they are solely invoked in the presence of other social agents (Elster, 1999: 141). He categorizes shame further as an internal interaction emotion, meaning that is based the agent’s belief about other agents, whilst contempt is external, insofar as it makes no claim on such a reference (Elster, 1999: 149). Shame is, then, invoked in the presence of others as a self-conscious negative evaluation of oneself, unlike guilt which is not contingent on the presence of others but works through, for example, the self-reflexive question “what would she have thought, had she knew that...” (Elster, 1999: 150). Here it seems that one could invoke Coleman’s (1990) description of the distribution of rights to influence and be influenced, because by passing on this right, one must also, although he did not consider it, attach some emotions to it – like in cases when one feels that one is violating the rela-

tional social contract, one has with another, although she might not be aware of it or even influenced by it, one may still feel guilty. Though not incompatible with Coleman's (1990) theory on social solidarity, it would, at least to some extent, violate the foundational assumption on rationality, see chapter III. Shame is invoked in situations when one's actions violate a social norm, unlike guilt which might even be invoked if one violates a private norm (Elster, 1999: 151). Guilt is local and related to a particular action which one has or has thought about carrying out, whereas shame is global and related to how one perceives one's character (Elster, 1999: 152). Elster (1999) argues on this point:

“In guilt, there is a correlation between what the agent feels and what others *would have felt* had they known what he did. In shame there is a causal connection between what others *actually feel* and what the agent feels.” (Elster, 1999: 153)

In that sense, guilt is related to what the agent believes that other social agents would think and because it is not, necessarily at least, invoked by the presence of others, the emotional intensity which accompanies it will be lower than the emotional intensity accompanying shame, insofar as this is causally related to how other social agents feel. Moreover, the action tendency of guilt is also, due to its local nature, to rectify who is wronged or to repair what is broken, unlike shame, here the broken will persist to be broken in the eyes of the ashamed agent (Elster, 1999: 153). Finally, unlike guilt, shame is not triggered by thoughts on intentionality or consequences within the ashamed agent herself (Elster, 1999: 154).

With this in mind, I now turn to Fehr and Gächter's (2000) definition of a social norm, because according to Elster (2005: 203) their definition resembles a quasi-moral norm, albeit with a moral normative *flavor*. The problem is, he argues, that social norms are often too weak to produce moral behavior, whilst moral norms centered upon normative claims are unconditional, whereas Fehr and Gächter (2000) implies that moral behavior is conditional on observing other social agents performing the act (Elster, 2005: 203). In addition, Elster (2005: 203) argues, based on consequentialistic moral reasoning, that moral behavior should be produced with regard to the consequences they have; rather than with regard to what others are doing. Social norms are enforced by shame, while according to the view of Fehr and Gächter (2000) they are enforced by a

sense of righteous anger invoked because the norm transgression comes to represent a deviation from a normative claim on reciprocity:

“A person is a strong reciprocator if she is willing to sacrifice resources (a) to be kind to those who are being kind (strong positive reciprocity) and (b) to punish those who are being unkind (strong negative reciprocity).” (Fehr, Fischbacher and Gächter, 2001: 3)

According to this perspective, at least, norm transgression is also accompanied with a belief on the transgressing agent's intentions. The agent sanctioning a norm violator, in Elster's (1999) perspective, does not do so, because she herself is moved by a righteous motive; rather she punish the violator, because he in some way has acted in a manner which violates the social norm. What seems, here, to be the focal matter is: on the one hand, a definitorial matter regarding how social norms are to be defined; but, on the other hand, it is also a matter of classifying the fairness norm. Elster's (2007; 2009a) distinction between social, quasi-moral, and moral norms enables him to differentiate between the nature of the different kinds of actions and motives which is to be categorized under and, as such, governed by each different norm. Depending on the action and its categorization, different norms will also be invoked, just as some actions although they do not invoke a moral normative expectation are still covered by social norms – the act of Christmas card sending, for example, binds the sender to continue sending Christmas card unless the receiver does not reciprocate, just as the receiver is expected to reciprocate the gesture (Elster, 2007: 354). This only to illustrate what seems to be Elster's (2005) argument here that all actions cannot nor should they be viewed as invoking moral like social norms, at least, insofar as not all actions are of a nature which allows them to be analytically classified as being governed by these moral like norms. Moreover, if an action is classified as belonging to the category of moral actions, then, one could argue that this is not, or perhaps rather should not, be governed by reciprocal concerns.

In chapter III, I argued that social norm theory has its intellectual roots in social contract theory and was introduced as a social convention enforced by the social nature of human beings associated with the socioemotional base of esteem/self-esteem. As such, social norms were originally introduced as a way of explaining social order. Social norms are related with the rules which govern social life, in the sense that they pre-

scribe or proscribe how to act in particular situations. Unlike the utilitarian approach to rationality, it is not governed by instrumental reasoning and, as such, it is non-instrumental in nature and not aimed at bringing about some goal besides itself. Unlike the Kantian perspective on rules, however, social norms cannot be reduced to matters related with morality. Moreover, unlike rational choice and moral judgment, social norms are not adhered to because of normative reasons – that is, the agent may not hold any strong reasons on this being the best or right way to act. Social norms, in other words, represents permissible and possible actions which the agent can choose to adapt her actions to. They govern social interaction as the expectations which accompanies certain actions, such as the Christmas card example discussed above, and when invoked they bind others as well as the agent self to a particular course of actions. Hence, when performing a certain action, social norms shapes one's expectations about how other social agents will react and respond.

Whereas morality is related with the agent's sense of what is right/wrong and good/bad; and practical rationality with the practical/technical issues related with how the agent should obtain her goals, social norms are related with the agent's understanding of social rules and, as such, it is related with the agent's ability to navigate in a social world – in the presence of other social agents. The agent may through technical and practical reasoning reach a decision on how she should realize her goal, but this decision needs to be in accordance with the agent's moral principles, just as she needs to take into account that not all actions are equally permissible from a social normative perspective. Whilst practical rationality is consequentialistic and related with the obtainment of some goal, moral reasoning and social norms are solely aimed at the action and not, necessarily at least, its consequences. In some way, these different factors symbolize the filters through which the action passes before being acted on. In a similar way, the social rational action is one which is in accordance with the agent's moral principles as well as social rules of a moral and non-moral nature. That is not, however, to argue that social norms are deliberately or strategically enacted, insofar as social norms may in some situations simply be enacted more or less automatically. On the other hand, if action is reduced enactment of principles this might reduce the possibility of agency, just this notion might underestimate the role played by acting agent, because no matter how one defines action it presupposes an acting agent – someone who choos-

es the action, who choose the action in accordance with the governing social rules. Of course, not all actions are instrumental in nature, just as they may not be produced in accordance with means-end reasoning – impregnated with a particular purpose – insofar as some actions are not internal constructions, but are acquired and learned. That is, some courses of actions are not produced as *unique* and *optimal* ways of bringing about some unique goal, rather they are learned patterns of actions and, as such, they are produced by other social agents, thus, they cannot be reduced to results of trial and error learning from personal experience, but represent how social agents occasionally mimics or adapts their behavioral pattern to that of other social agents. In that sense, just as not all actions are instrumental, not all courses of actions are internally produced, some are acquired or learned. Hence, these pieces of action still needs to be collected, just as they needs to be applied in an appropriate manner and this can only be done by a human being.

### **8.3 The Need to Belong and Social Norms**

In the previous chapters, I have discussed the nature of social rationality and argued that social rationality follows as a necessary consequence of the relational structuring of social life – that is, it follows from the fact that taking part in social life implies an exchange of rights and obligations. In an organizational context, social rationality is important, insofar as it contributes to the continuation of the social relations – vertical as well as horizontal. Furthermore, it was argued that the social optimal choice was that which maintained the social relations, because these represent the productive foundation of the organization, regardless of how the work in the organization is organized. The explanation proposed so far, is, at least to some extent, functional in nature, in the sense that the focus is on how social rationality serves a systemic function in maintaining the social functionality of the organization. It should, however, be noticed that while social rationality contributes to the functionality of the organization this is primarily a by-product, insofar as the social norms governing the social relations may not have been purposively designed to serve this function. Indeed, some relational norms may even have adverse and counterproductive results as argued by Elster in chapter III.

The aim in the present chapter is not; however, to propose such a functional theory, rather the purpose is to discuss the social dynamics of differentiation. In chapter III, I argued that social agents may act morally towards other social agents due to a cost/benefits analysis, esteem/self-esteem concerns, sympathy, or a self-constituted obligation. In the section above, I demonstrated that a similar reasoning can be invoked when explaining commitment to social norms. The Hobbes-Coleman perspective argues that social agents adheres to social norms due to a cost/benefits analysis, Elster's Humean inspired perspective focused on concerns regarding esteem/self-esteem, while the Bicchieri-Fehr-Gächter approach argued that it was due to the possibility of external sanctioning. From a metatheoretical perspective, then, the underlying logic across the models is egoistic in nature and brings with it negative associations.

From the perspective pursued so far in this thesis, social agents adhere to these social norms, because it enables them to maintain existing social relations. A similar understanding is proposed by Elster (1989b; 1999; 2007; 2009a), who views ostracism as an important form of social punishment. Loss of belongingness may, in other words, follow as a consequence of norm transgression either in the sense that one retracts from the company of others due to shame; or in the sense that others keep their distance. While it might be problematic to argue that individuals want to fit in and to be accepted – it might be less problematic to argue that individuals want to belong and to be part of something which extent their individual life. Hence, the aim of this section is to understand this social need and its implications for strategies on self-differentiation.

In sociology, Durkheim (1915/2008: 209, 226, 230) addressed the existence of collective sentiments which, when internalized, works through the agent and generates the foundation of moral judgment, just as he discusses the transcendent state of *collective effervescence* when the individual loses herself in a collective state of mind. Recently, another of his concepts, related to this, has been discussed in sociology, *Homo Duplex* (see, for example, Fish, 2013). According to Durkheim (1915/2008: 15-16, 223, 438), this concept describes a particular way of being – a social being – split between the material and individualistic world and a collective, non-material, social world from which morality originates. In a way, then, this resembles Boudon's (2001) conception of the impartial spectator discussed in chapter III.

In theories on work motivation, Vroom (1964/1995: 48) viewed interpersonal relationships, horizontally and vertically, as one of the bases of work-motivation, while Herzberg, Mausner, and Snyderman (1959/2010: 93, 115) viewed it as a rather insignificant hygiene factor. In his existential psychology, Abraham H. Maslow (1962/2011: 26, 31-32, 38, 168) distinguishes between the *need-deficit motivated* agent who lacks safety, belongingness, love, respect – all needs which can only be satisfied by other social agents, thus she is dependent upon these others, and as such she is not in control of her own fate, but must constantly be sensitive to the wishes and rules of others when she desires to obtain their approval and affection; and the *self-actualizing* autonomous and authentic agent who is self-contained, governed by inner laws and desires which springs from an autonomous self. A need-deficit motivated social agent solely perceives other individuals as means to an end – her satisfaction – whereas the authentic agent is capable of seeing these other social agents as true, unique, and complicated individuals (Maslow, 1962/2011: 33).<sup>37</sup> Although Maslow's (1962/2011) perspective is highly influenced by existential philosophy and Buddhism, it illustrates that to reach the level of authenticity discussed by, for example, Deci and Ryan (1985) in their theory on human motivation, the social agent would need to have her social needs satisfied – implying that solidarity and social relationships matter and not just in a negative sense as seems to have been emphasized above.<sup>38</sup> Additionally, although Deci and Ryan (1985) in their

---

<sup>37</sup> The perspective seems, at least to some extent, closely related to that proposed by Robert W. White (1959). According to White (1959: 297, 318, 321) when wanting to understand human growth, one should focus on what he called *competence* – that is, the individual's ability to interact with her environment, just as one should focus on the individual's motivation to do so, and even though external stimuli is important the motivation behind competence comes from within in opposition to drive theory in which the environment provides stimuli to satisfy the organism's immediate need.

<sup>38</sup> To reach the level of authenticity in self-determination theory, the agent needs to feel competent and autonomous. *Competence* may be understood as the agent's sense of her own ability or mastery of the task at hand and as such it is also linked to the agent's sense of efficiency – that is, with regard to monetary extrinsic rewards the sense of competence follows from meeting the criteria and receiving the reward, while competence is more closely related to intrinsic motivation which follows more or less directly of the sense of mastery – of being efficient in solving an optimal challenge. Being impelled by a need for competence, then, implies engaging with the world in a positive manner – a manner in which the agent desires to learn, develop, grow, and simply to master (Deci and Ryan, 1985: 58, 123-124). Competence without autonomy, however, is not motivating, thus praise (or other kinds of external stimuli) may in some cases be motivating if it adds to agent's self-perceived level of competence, and if it is not perceived as controlling. *Autonomy* refers to the sense of self-determination and the need for internal causality – that is, the choice or action must by the agent be perceived as being a result of her own will. The agent, in other words, pursues a course of action which she finds to be right and not because she is forced to it and neither is the action produced with different motive than the one simply expressed in the intention with which the action was produced. It is, nevertheless, autonomy within limits meaning that limits should be provided which supports this sense of autonomy – these autonomy supporting limits must, of course, not be accompanied by a tacit requirement or a hidden agenda because if so they might have ad-

early work did not explicate the social dimension, they later emphasized it when they argued that social agents, even at the state of authenticity, do need to feel a sense of belongingness or relatedness (Deci and Ryan, 2000). These basic social needs, the authors, associate with their theory on organismic integration which refers to the concept of competence (the agent's perceived sense of mastery, for a further conceptualization see the footnote) describing how the agent engage and interact with the context in a proactive manner, in a continuous move towards a sense of greater coherence, consistence – or simply inner harmony. Human behavior and experience, in other words, is viewed as a dialectic process between the individual agent and the context – the interaction between the individual and the context, can either be positive and nurturing or antagonistic, thus in case of the former, a synthesis state occurs in which the agent's proactive tendencies are supported in manner allowing them to develop and the individual to grow, whereas in case of the latter, a synthesis state does not occur and this might undermine the agent's intrinsic motivation due to internal conflicts (Deci and Ryan, 2000: 229). The synthesis state, they call *integration*, a state characterized by the agent's free and proactive adaption to the values and regulations of the context in manner which allows her to integrate these with her own values and rules and eventually this can be seen as the agent's autonomous acceptance of responsibility; while in the opposite case, when integration does not occur, the authors call this state *introjection* and this be viewed as a state of control and coercion in which the rules are perceived as demands which demeans, and agency is reduced to purely instrumental reasoning (Deci and Ryan, 2000: 236). As also argued by Maslow (1965/2011) above, and Plato and Kant in chapter III, only at the state of authenticity can social agents engage in pure relationships with one another, because only at this state do they perceive other social agents as other human beings and not just as means to satisfy their private needs – that is, autonomy is not equal to independence and neither is it contrary to dependence (Deci and Ryan, 2000: 242). Consider the following relationship:

---

verse consequences, undermining the sense of autonomy and the intrinsic motivation following from this (Deci and Ryan, 1985: 149, 153, 158).

**Table 2: The logic of social relatedness and autonomy.**

	<b>Autonomy</b>	<b>Non-autonomy</b>
<b>Independence</b>	Independent by choice	Independent not by choice
<b>Dependence</b>	Dependent by choice	Dependent not by choice

The agent who chooses independence by will, belongs to the upper left corner of the model, she is not forced in any way nor have she been *mistreated* in a way which might make her choose this way of living. Deci and Ryan (2000) interpret autonomy as free of external coercion/influence in a similar manner as Kant, see chapter III. In chapter III, I argued that true autonomous motivation emerges from within, from the will, just like true self-esteem springs from a true idea of self-worth, uninfluenced and independent of the opinions of others. In order to distinguish between social relational autonomy and the autonomy implied in Kantian agency, one may distinguish between cognitive autonomy and social relational autonomy. In accordance with Kantian agency, the agent should retract herself from the influence of the external world – from the extrinsic motivational factors – and find the reason and motivation within herself in accordance with her will, thus it entails cognitive autonomy. In a similar manner, a Kantian agent commits herself to take on the obligation which stems from being the cause of her actions and, as such, the social tie between individual agents originates from a true sense of responsibility, from a will to be responsible, from an autonomous choice in which the agent chooses to choose the other – a choice which is not based on what the agent hopes to achieve, insofar as this would be in opposition to the principle stating that one must never treat other agents as means to one’s end. As argued in chapter III, among other things, it is this social commitment which distinguishes Kant’s perspective from Plato’s, in the sense that Plato did not recognize the outward directed nature of social life, but focused solely on the inward directed responsibility associated with governing oneself. The true Kantian agent chooses to choose the other and commits herself to respect the other individual agent’s rights as another agent belonging to the kingdom of reason. Nevertheless, from this it seems to follow that the Kantian agent who asks nothing in return for her true commitment can never be dependent upon another agent – only in a moral sense. In that sense, the agent commits herself to respect the moral rights of oth-

ers and expects them to do the same, but she is not dependent and neither is she truly independent, because this would imply that she did not feel any outward directed responsibility. A true Kantian agent is, in other words, independent by choice in the sense that that this state enables her to gain cognitive as well as social autonomy, and dependent by choice in the sense that she acknowledges the responsibility following from being the cause, from acting in a social world.

The category, the non-autonomous state of independence, at the upper right corner of the table, is where the agent for some reason sees herself as not being able to relate or rely on other individuals – this could be where Elster's (2009a) ostracized agents would be placed. At the bottom left corner of the table, one finds the state of autonomy combined with dependence which is the one emphasized by Maslow (1965/2011) and Deci and Ryan (2000). Finally, at the bottom right corner, one finds the non-autonomous dependent agent, who not by choice is dependent on other social agents, here one could, for example, place the agent described by utilitarian philosophers, and the agents described by Elster (1989; 1999; 2007; 2009a) and Coleman (1990). The social agents described by Smith and Boudon (2001), who accepts their need for belongingness, but not always in the positive, perhaps too positive, manner described by Deci and Ryan (2000), hence they could be centered somewhere between autonomous and non-autonomous dependency. The problem is, of course, as also seems to be emphasized by Elster (2007), that this need to belong which propels agents to engage in social relations also can lead to introjection, because when social relatedness is contingent upon showing solidarity through norm adherence, then the social agent's need for belongingness might lead her to adopt values which she cannot identify herself with, just as their sense of self-esteem may also be tied up with this. That is, the social world describe by Elster (2007) seems to be a world in which other social agents capitalize on this social agent's need for relatedness by threats of ostracism and contingent relatedness. Furthermore, the theory also seems to direct the attention to the potential problems of reducing this to a matter of instrumental reasoning or rational calculation as Coleman (1990) does, insofar as this seems to understate the psychological necessity of relatedness. The agent who solely shows solidarity out of instrumental motivation belongs to the bottom right corner of the model.

This categorization can, of course, primarily be viewed as simple speculations and reflections and has, as such, solely as its aim to enable a practical understanding of the table. Notwithstanding this, the table directs the attention towards a problem in the classic philosophy of rationality related with the understanding relatedness. As I argued above, the Kantian agents are solely bound together through a self-commitment to an outward directed moral responsible. The Humean agents are bound to one another through the socioemotional structuring of esteem/self-esteem, while the late Smith views economic rationality as being associated with relatedness through exchange and the early Smith recognizes relatedness through true non-contingent sympathy. In the Hobbes-Coleman approach, relatedness is associated with influence – the fact that acting in a social world entails the production of externalities and that the agent is never truly autonomous or not influenced. Relatedness, then, is either described in a moral sense as an outward-directed responsibility; as something which emerges from exchanges or externalities; or as something emerging from the constellation of human nature as socioemotionally connected. Thus, in addition to the characteristics provided of the instrumental, the relational, and the moral approaches to organizational fairness, it should be recognized that these models addresses different understandings of social relatedness – why they emerge and what they entail – emerging from their intellectual roots. Moreover, it might also be argued that in accordance with the framework developed in this section, social agents are responsible to one another and dependent upon one another due to material or emotional factors. In that sense, while the Kantian perspective comes close to the understanding of relatedness associated with social rationality pursued in this thesis, it does not entirely capture the pragmatic necessity of relatedness implied, hence in the following I will seek to identify what is actually implied and what this entails in sense conceptual commitment to a particular understanding of human relatedness.

### **8.3.1 In Search of an Understanding of the Pragmatic Necessity of Relatedness**

In psychology and moral psychology the concept of relatedness has also recently been emphasized. Roy F. Baumeister and Mark R. Leary (1995: 499, 500) identified, in their classic article, the need to belong as a basic psychological drive – the authors de-

defined as the need for frequent and stable long-term interaction with few, but significant others. Social agents are, as a result of this need, motivated to form a minimum of stable social relationships which needs to be maintained, positive emotional bonds will follow from the establishment of these relations, and beyond this minimum, additional social relationship will bring decreasing satisfaction (Baumeister and Leary, 1995: 500, 505, 517). If this minimum of stable social relations with significant others is not met, the social agent will experience different psychological consequences, for example, loneliness or isolation – the requirement regarding stable relationships states that relationships without emotional ties, bonds of caring, or frequent interaction may not bring the same level of satisfaction and can lead to a state of partial deprivation (Baumeister and Leary, 1995: 511. 513. 515).

Haidt, Patrick Seder and Seling Kesebir (2008: S135) call this the first and basic step towards recognition of the social need. In their article, the authors present the *hive hypothesis* which states that social agents has a need to, like argued by Durkheim (1915/2005), lose their sense of self in a kind of transcendent collective experience, to reach the highest state of human well-being (Haidt, Seder and Kesebir, 2008: S136). More recently, Haidt (2012: 194, 223) have provided a more elaborate statement of the hive hypothesis – here he argues, based on a group selection approach, that human nature has a *groupish* overlay which he explain as a product of human evolution, because the groups that were able survive were those which was most coherent – that is, promoted solidarity with the group's needs, hence as a byproduct of this, social agents are equipped with what he calls a *hive switch* which enables them to, temporarily and under certain conditions, transcend self-interest and enter a state of collectiveness. The metaphor of the switch is applied, because it is only conditionally *flipped* on. The hive switch is biologically linked with the production of oxytocin, a hormone which, for example, stimulates milk letdown in non-human mammals, just as a study have shown that if oxytocin is sprayed into the nose of players in a trust game they will choose higher levels of transfers than the placebo group – that is, oxytocin enhances trust in interpersonal interactions; and neurologically linked with the mirror neurons which, for example, explains empathy (Haidt, 2012: 233, 236; Kosfeld, Heinrichs, Zak, Fischbacher, and Fehr, 2005: 674).

A study by Tania Singer, Ben Seymour, John O’Doherty, Klaas Stephan, Raymond Dolan, and Chris Frith (2006) demonstrates how like/dislike bias influences players’ behavior in a sequential prisoners’ dilemma. The respondents played against a member from the research team who could either play fairly or unfairly. In the second part of the experiment, the member from the research team was exposed to mild electric shocks through his/her hand which was clearly visible to the other player (Singer, Seymour, O’Doherty, Stephan, Dolan, and Frith, 2006: 466). The study showed that the respondents emphasized more with a fair player than an unfair – in fact, neuro-scanning showed that seeing the unfair player *suffer* activated the part of the brain commonly associated with revenge (Singer, Seymour, O’Doherty, Stephan, Dolan, and Frith, 2006: 467). In that sense, the study supports social preference theory, in the sense that social agents evaluate gains of fair players positively and gains of unfair players negatively. In other words, the hive switch explains, biologically and neurologically, why social agents, occasionally and conditionally, act in a non-selfish manner, because it promotes within group solidarity and personal well-being. The need to belong, then, seems to be positively related with human well-being. Just as it seems to explain group solidarity as being based on a different motivational basis than concerns about esteem and self-esteem, insofar as the driver here seems to a desire for belongingness and not just a desire for social approval – that is, social agents are not just motivated by the impartial spectator, they are also motivated by a desire to belong. This can, of course, be related to the vertical dimension, in the sense that organizations which succeed in activating the hive switch will be able to foster loyalty. It may, however, also be related with horizontal relations between colleagues, insofar as it seems to demonstrate that social agents are not just motivated by positional or monetary concerns, they are also motivated by belongingness.

The need to belong has also been shown to be related with contextual and interpersonal sensitivity. In a study by Cynthia Pickett, Wendi Gardner, and Megan Knowles (2004), the authors studied how the need to belong interacted with the social monitoring system – the latter should be conceived as a way of explaining social agents’ sensitivity to social stimuli – subtle cues such as vocal tone, facial expressions, etc. providing information about approval/disapproval (Pickett, Gardner, and Knowles, 2004: 1096). Initially, the respondents in the study were asked to introduce themselves to one another

and then they were lead separate computers and asked to enter who they wanted to team up with. One group of respondents were told that none of the desired partners had wished to team up with them (social-exclusion), while another group was told that the number of teams had been exceeded (non-social exclusion) (Pickett, Gardner, and Knowles, 2004: 1098). The respondents were, then, asked to answer some questions which enabled the authors to assess the individual respondent's need to belong, the respondents were then exposed to different images, video, and sounds related with social interaction. The study demonstrated that the respondents' degree of need to belong was positively correlated with, for example, the respondents' sensitivity to and accuracy of reading facial expressions and identifying vocal tone differences (Pickett, Gardner, and Knowles, 2004: 1104). In a more recent study, Namkje Koudenburg, Tom Postmes, and Ernestine Gordijn (2013) studied how the need to belong interacted with social agents' sensitivity to social norms. The point of departure for the study was how social norms often are maintained by more subtle means than punishment – more specifically, the authors focus on the meaning of conversational silence as a sign of norm violation, which although subtle can be viewed by the norm transgressor as a sanction (Koudenburg, Postmes, and Gordijn, 2013: 225, 226). The respondents were first asked to fill out a questionnaire thereby allowing the authors to assess the individual respondent's need to belong. The respondents, then, watches a video of a conversation between two individuals and they were asked to identify with the person who would turn out to be the norm-transgressor (Koudenburg, Postmes, and Gordijn, 2013: 228). The study demonstrated that if the norm violation was followed by conversational silence, the respondents perceived the non-norm violator's attitude as more threatening compared to the opposite, if the conversation just continued (Koudenburg, Postmes, and Gordijn, 2013: 225, 229). Moreover, respondents with a strong need to belong showed higher levels of motivation to conform to the norm if the norm-transgression was followed by conversational silence than if not (Koudenburg, Postmes, and Gordijn, 2013: 225, 230). Combined, these two studies seems to support Elster's (1989; 2007; 2009a) claim that social agents are motivated to norm following due to interpersonal concerns, however, as these two studies seems to argue, besides the concerns about esteem and self-esteem, the social agents are also motivated to conform due to a need to belong.

In the recent literature on social ostracisms, research have, for example, found that experiences of social ostracism (even from non-human actors) undermines social agents' sense of belongingness and self-esteem (Zadro, Williams, and Richardson, 2004: 563), increases aggression (Twenge, Baumeister, Tice, and Stucke, 2001: 1060), and reduces pro-social (helping) behavior (Bartels, Baumeister, Ciarocco, DeWall, and Twenge, 2007: 58). Brock Bastian and Nick Haslam (2010) studied how social ostracism might influence social agents' perception and attribution of human values to self and others. The authors distinguishes between two kinds of dehumanization; *human uniqueness* associated attributes distinguishing humans from non-humans (e.g., civility, morality, higher cognition, etc.); and *human nature*, associated with attributes of shared fundamental features of humanity (e.g., emotionality, agency, etc.) (Bastian and Haslam, 2010: 167). When the former is denied, social agents are perceived as irrational, likened to animals, etc., whereas when the latter is denied, social agents are likened to machines or objects (Bastian and Haslam, 2010: 107). The latter state of interaction is associated with what Fiske (1991: 19) calls the *Null Relationship* which he views as the basic point of reference to be distinguished from relations structured emotionally as well as cognitive. Bastian and Haslam's (2010: 108, 112) study shows that targets of social ostracism perceive themselves and others as less human compared to neutral or cases of social inclusion – that is, they deny themselves and others traits commonly associated with human nature, just as they also seems to believe that others view them as less human. In other words, the targets of social ostracism view themselves as excluded from humanity, from the human category – and the detrimental psychological consequence is a diminished sense of humanity (Bastian and Haslam, 2010: 112). In a more recent study by Bastian, Jolanda Jetten, Hannah Chen, Helena Radke, Jessica Harding, and Fabio Fasoli (2013) the authors study the self-perceptions of norm transgressors. The authors find that these agents tend to dehumanize themselves and view their actions as immoral (Bastian, Jetten, Chen, Radke, Harding, and Fasoli, 2013: 161). The self-dehumanization occurs regardless of global character evaluations (self-esteem) and emotional reactions (mood) – that is, it occurs regardless of whether the social agents perceive themselves as bad or feel bad. In general, their study also shows that social agents, who as a consequence of their actions dehumanize themselves, tend to engage in pro-social behavior to rectify or perhaps restore their sense of humanity through initiat-

ing positive relations with others or performing self-sacrificing actions (Bastian, Jetten, Chen, Radke, Harding, and Fasoli, 2013: 164). In relation to Elster's (1999) theory, one might speculate that this self-dehumanizing effect which occurs is associated with shame, however, the moral engagement which follows as a result of this may, according to Elster's (1999) distinction, be associated with feelings of guilt. The authors study also showed that these self-evaluations occurred in social agents who themselves had ostracized others, and as such violated an interpersonal norm. Hence, in addition to Elster's (1999) logic of sanctions, one could add the non-material costs to the ostracizer, in the sense that this also seems to be a psychologically painful act to perform which might have consequences to the ostracizer herself, thus social ostracism is painful to the ostracizer and the ostracized. Another interesting aspect of this study is that these self-evaluations was more related to perceived immorality than belonging – a fact which seems to support the above point on its relation with the social emotion, guilt (Bastian, Jetten, Chen, Radke, Harding, and Fasoli, 2013: 165). In other words, by performing such norm transgressing behavior the social agent comes to dehumanize herself due to perceived immoral behavior which implies that self-dehumanizing does not, necessarily, lead to moral disengagement; rather it seems that the effect occurs as a matter of the opposite.

From what has been argued so far in this section, it seems to follow that social agents have a need to belong. This need to belong, however, is not related with a self-commitment to a particular obligation and neither is it related with obtaining material goods – although, in some cases it might be associated with obtaining socioemotional confirmation. Rather, this need to belong is related with being human – with that which makes humans perceive themselves as humans – that is, in the eyes of agent self, being ostracized, being independent not by choice, comes to symbolize unworthiness and as being stripped of one's humanity and uniqueness, one no longer perceives oneself as being human or being worthy of the company of others. Anomie, as discussed by Merton in chapter V, does not solely emerge from being denied that which one perceives as being within one's right to claim, it also follows from being ostracized, from being denied one's humanity. The need to belong may also explain why social agents adheres to social norms, insofar as norm transgression is associated with acting against the social rules and this may, eventually, lead to a loss of relatedness. This loss of relatedness is,

as argued above, painful because it symbolizes a loss of humanity, but also because it goes against the basic human need. While the dimension related with loss of humanity may be socially constructed, the need to belong is also *hardwired* in the human biology, in the brain and the signals it transmits – as Haidt (2012) argues, humans are wired to connect. As argued by Haidt, inspired by Durkheim, it is this conditional hive switch, the fact that humans have the ability to transcend their private interests and material life which makes morality possible. From this perspective, the need to belong enables social rationality, just as it impregnate it with a sense of necessity, in the sense social rationality becomes a source of securing relatedness, just as it is enabled by this. Thus, without this need to belong, social rationality would not be necessary and without social rationality social relations could not be maintained. Without social rationality, the social world would be reduced to a Hobbesian state of war; however without a need to belong social relations would be reduced to matters of costs/benefits analysis. The Hobbesian analysis and those inspired by his perspective (see, for example, Pinker, 2011) might be right in the first part of the thesis, but the reason is not, entirely at least, to secure material gains and a stable society, rather maintaining these relations becomes a goal in themselves as it seems to be the main conclusion one can draw from the research discussed above.

From the perspective developed in this section, then, acting against the governing social norms may be painful to the norm transgressor – she may be punished by those around her or she may punish herself. In some sense, acting against the governing social norms is acting against one's interests, in the sense that norm transgression may influence one's psychological health. If it is irrational to act against one's self-interests, then, acting against the social norms is irrational, insofar as it may lead to a loss of belongingness. Of course, the agent still needs to evaluate the norms in relation to her own moral principles, because not all social norms may be in accordance with her moral principles. The individual who chooses to differentiate herself, nevertheless, sets herself above the group and by pursuing her own ends she is acting in her private interest and not that of the group. The aim here is not to argue that social agents should subdue themselves unconditionally to the group's norms. Rather, the aim is understand how differentiation, in some situations, may be opposite to the agent's interests. Because the agent differentiating herself is setting herself above or aside from the group and, as

such, she is excluding herself. By choosing to differentiate herself the agent is transmitting a particular signal to her colleagues and to herself. From this perspective, the social bonds existing between the agents are not threatened, as argued by Lawler (2003) in chapter II, by differentiation of vertical relations or outcome differences and the fairness issues which might be invoked. Rather, the bounds are threatened by the very act of self-differentiation and matters related with fairness in a social rational sense. What rip apart social bonds, then, may not necessarily be matters related with outcome fairness and egoistic motivation, rather it may be associated with the actions which precedes these outcome states – in the sense that outcome states follows as by-products and, as such, no causal relation can be established, see chapter VI. Thus, differentiation in itself as a state may not be against social rationality, but the actions needed to bring this state about may be. That is not, however, to argue that social agents should adhere to norms similar to ones described in chapter II, norms which can be viewed as being primarily egoistic in nature and related with protecting others from looking bad and keeping down those who are better performers. The aim is, solely, to argue that better performers solely emerge as such, because they choose to differentiate themselves and by choosing this they may be acting against their own need to belong, they may ostracize themselves from the group and in the long run this may undermine the social foundation of the organization's productive core.

### **8.5 Tentative Conclusions**

The aim of this chapter has been to discuss the dynamics of self-differentiation from a socially rational perspective. Self-differentiation related with choosing to differentiate oneself from one's peers may be psychologically painful, insofar as this act of self-differentiation may lead to one's ostracism from the collective either by others or by oneself. If self-differentiation leads to social exclusion, it may be viewed as being against the psychological need to belong. Hence, self-differentiation may come with a price.

Self-differentiation may, in other words, in some situations undermine the social ties which bind the agents to one another, in the sense it may be viewed as being in opposition to the governing social norms – just as it might be viewed as antisocial behav-

ior and lack of respect for the social norms. As some studies show, the need to belong also influences the agent's propensity to norm following, because they come to associate norm transgression with loss of relatedness.

This need to belong follows from a psychological need to belong – a need to be part of something which extends one's private and material interests, thus it might be viewed as the foundation of morality. The need deficit agent who lacks this sense of relatedness may find herself in a state psychological anxiety, in the sense that this lack of belongingness might make the agent perceive her and others as less humane and, as such, the individual may retract herself from the company of others. The agent who retracts herself may not do so entirely by choice. Some agents may, of course, choose to retract themselves from social life. Nevertheless, the need deficit individual seeks the company of others in order to reduce this psychological stress and, as such, she perceives these other social agents purely as being a means to her own end. The non-need deficit agent, on the other hand, does not perceive other social agents as being means to her own end, rather she perceives them as goals in themselves and she chooses to choose these individuals and perceives them as individuals on their own terms. The agent at this psychological state, chooses to be dependent, just as she is cognitive autonomous, in the sense that her actions springs from her will and not from some extrinsic factors. Only at this psychological state does the agent live an authentic life, only at this state does the agent engage with the world in an autonomous manner. But in order to reach this state of authenticity and self-realization, the agent has to have her social needs satisfied. Social rationality, in that sense, may be viewed a way of securing this, because the socially rational agent is able to secure her relatedness by acting towards the other in a manner which signals respects. That is not, however, to argue that social agents should subdue themselves unconditionally to the group's norms or the desires of other agents, insofar as this would be in opposition with the principles of an autonomous social life, just as it would be in opposition to the principle of self-governance emphasized by, among others, Plato and Kant, see chapter III. Just as not all norms may be in accordance with the agent's moral principles and some norms may be counterproductive, in the sense that they solely protects the interests of some.

Social rationality may, then, be viewed as a principle of action which not only secures the continuation of social relations, it also secures that social agents can obtain this sense of relatedness. The perspective described here deviates from the conceptualization of social relations inherent to the classic philosophy on rationality, insofar as social agents do not, solely at least, seek relatedness in order to obtain either material or socioemotional confirmation and neither can this sense of relatedness be reduced to sense of felt obligation following from sympathy or self-commitment. Only the non-need deficit social agent can truly commit herself and feel sympathy, whereas the need deficit agent reduces her social interaction to a matter instrumental/strategic interaction, in order to satisfy her egoistic need. Hence, non-instrumental and autonomous interaction presumes non-need deficit agents – because only then, can moral agency in Smith's and Kant's sense be realized.

This description may also be viewed in opposition to some of the dominant perspectives on social norms, insofar as the motive behind norm following cannot be reduced to fear of external sanctions or a costs/benefits analysis. Elster's theory on social norms provides a similar explanation, because he argues that agents primarily adhere to social norms to escape the shame accompanying norm transgression. The feeling of shame may be associated with many of the features already described. It might however, be argued that this focus on the negative dimension of norm following does not, explicitly at least, capture the psychological need which might also motivate social agents to norm following.

On the one hand, then, from a motivational perspective it might be argued that acts of differentiation may undermine the agent's motivation, insofar as what might be gained in terms of outcome, for example, external symbols (monetary or non-monetary) is obtained at the expense of relatedness, implying that the agent in the long run might experience psychological anxiety related with this lack of relatedness. On the other hand, from a social perspective, it might be argued that in the long run the social relations are not undermined from the outside by factors associated with compensatory differentiation and differences, rather they are undermined from within, due to self-differentiation the agents may ostracize themselves and others and as a consequence

transform social relatedness into a matter of costs and benefits – hence, realize the Hobbesian state of war.

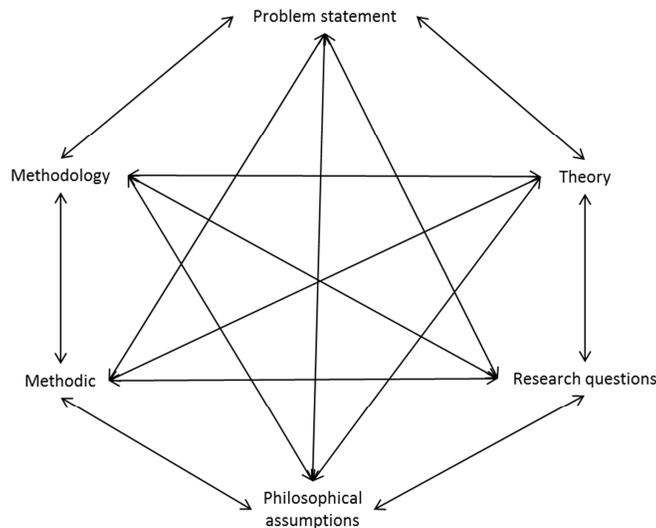
# Chapter IX

## Methodological Reflections

### 9.1 Introduction

The aim of this chapter is to clarify and discuss the methodological foundation of the empirical analysis in the next chapter. Throughout this chapter, I will distinguish between *methodology* related with the theoretical foundation of the chosen research methods and *methodic* related with the practical application of these in the present thesis (Creswell, 2003: 5). Furthermore, I will also discuss the relation between the chosen methodologies, their practical application, and the philosophical foundation of this thesis, see the figure below. Following upon this, a discussion of the overall quality of research program will be made. The chapter will end with a discussion of the methodological and empirical delimitations on which the research design is centered.

Figure 35: The research process illustrated through an argumentation model (inspired by Niglas, 2001).



The figure, above, illustrates the research process through the connections between different argumentations – that is, it is centered on the assumption that internal reliability is centered on the coherence of the different arguments regarding the different choices inherent to the research process. In chapter I, the philosophical assumptions of this the-

sis was explicated – here it was argued that the aim of science is to identify the link between the intransitive and transitive dimension of reality, the transitive dimension being constituted by pre-existing theories and the purpose of the research program. In that sense, the research questions are centered upon the problem statement, although these pre-existing theories go into the construction of both the problem statement and the research questions. The methodological question, then, becomes a matter of identifying the methods which enables one to collect data on these generative mechanism – that is, it becomes a matter of collecting data on the state of knowledge which enable the existence of the social phenomenon of interest. In other words; (1) what knowledge claims are made – what kind of data is needed? (2) What kind of methodology will enable one to collect the type of data needed? (3) Which method(s) of data collection will enable one to collect the *right* data and the data needed? (Creswell, 2003: 5; Hammersley, 2013: 4; Danermark, Ekström, Jakobsen, and Karlsson, 2002: 152). These questions explicate the transformation of the knowledge question from a question about which knowledge enables the production of reality to a question about which kind of knowledge must be collected in order to tap into this knowledge.

## **9.2 Research Design**

The aim of the present research design is twofold, in the sense that, on the one hand, the aim is to enable an inquiry into the underlying structures of fairness judgments, just as it, on the other hand, is to explore the relation between fairness, horizontal social norms, and relational influence (see also chapter I). More specifically, the aim is explore the relation between compensatory differences, social norms on horizontal differentiation, fairness, and relational influence. To capture the underlying structures of these judgments I will draw on a quantitative and qualitative frame. The aim of the quantitative frame is to study the patterns between the contextual frame, meta-judgments, and the primary judgments on fairness and relational influence. The aim of the qualitative study is to supplement the quantitative study, just as the qualitative study enables an *in-depth* study of the underlying structures (Creswell, 2003: 21). The quantitative study enables one to identify patterns in judgments across the respondents, while the qualitative study, enables a study the interpretive pattern which go into these judgments – the associative

patterns and cognitive frames (Lieber, 2009: 219). That is, the standardization of responses in the quantitative study allows direct comparison across respondents, whereas the non-standardization in the qualitative study enables a more thorough interpretation of the cognitive frames lying behind these judgments (Kvale and Brinkmann, 2009: 88, 117). Studying patterns implies studying consistency and variance across respondents, but these become more interesting when combined with a study of interpretive patterns (Maxwell and Mittapalli, 2010: 159; Danermark, Ekström, Jakobsen, and Karlsson, 2002: 165; Sayer, 243).

The approach described here follows Martyn Hammersley's (2013) approach to mixed studies. Hammersley (2013: 11) calls his approach methodological eclecticism and distinguish between three approaches: (1) *triangulation* aiming at testing the results of the two methodologies against each other; (2) *facilitation* aiming at using the two methodologies to generate the foundation of one another, for example, using one method to generate the hypothesis to be tested in another; and (3) *complementarity* aiming at using the two methodologies to support one another (Hammersley, 2013: 11-12). Furthermore, in the third approach, the researcher aims at integrating the two types of data into an overall interpretation of the results (Creswell, 2003: 16). In the two first approaches, the processes of data collection are somewhat *dependent* on one another – that is, triangulation implies collecting data simultaneously (*concurrently* data collection), while facilitation implies collecting data sequentially (Creswell, Clark, Gutmann, and Hanson, 2003: 215; Morgan, 1998: 366-367). In the third approach, the process of data collection associated with the two methodologies implies *independence*, in the sense that the order of data collection has no impact on the application of the methodology.

The approach applied here is inspired by Hammersley's (2013) third approach. This should not, however, be understood in a naïve sense – as building on the postulate that applying two methodological approaches enables one to surpass the limits inherent to them. On the contrary, the application of a mixed methodological approach introduces a new set of problematic aspects which need to be taken into consideration (Hammersley, 2013: 13). Nevertheless, a mixed methodological approach allows one to add another dimension to the research subject – that is, like the application of different theories allows one to assess different dimensions of some phenomenon, the application of

different methodologies allows one to study different dimensions of the phenomenon. In that sense, the aim here is not, as such, to use one methodological approach to cover the gaps of another; rather the aim is simply to understand two dimensions of the research subject – patterns in judgments across respondents and cognitive patterns of judgments (see also the discussion of mechanisms in chapter I). Additionally, neither of the two dimensions can be separated from the theoretical dimension, insofar as the theoretical dimension adds the interpretative frame within which the combined results of both studies are changed from mere information on patterns to actual categorized knowledge (see the discussion of the transitive dimension of science in chapter I).

In accordance with the philosophical perspective presented in chapter I, the two kinds of data generate the concrete/non-abstract level of the analysis, insofar as they combined/integrated generates the phenomenon which needs to be explained with a reference to more general/abstract existing mechanism. The role of the theories is two-fold, insofar as they, on the one hand, generate the assumptions on which the quantitative and qualitative studies are based, just as they, on the other hand, goes into the analysis of their results. The results of the study are, in other words, produced in a dialectic process between the results produced by the quantitative study, the qualitative study, and theory. In that sense, the *vertical* dialectical process between the results of the respective empirical studies and the *horizontal* dialectical process enables the production of results which are not only anchored in theory, but also in empirically in two different studies, see also below (Clark, Creswell, Green, and Shope, 2010: 380). As such, it might be argued that what enables the interpretation of the across respondents pattern is not only theory, but also the cognitive patterns. Viewing the across respondents patterns as the result, what enables the search for a best possible explanation is not just theory, but also an inquiry into the cognitive structures which might produce the foundation of these patterns – that is, the state of knowledge implied which goes into the construction of these results (Maxwell and Mittapalli, 2010: 157). The study may, then, be viewed as explanatory, in the sense that the quantitative results will first be analyzed and, based on this, the results which needs to be followed up – further explained – by qualitative data will be identified (Clark, Creswell, Green, and Shope, 2010: 372, 374). This does not, as already explained, imply that the data collected is dependent. Rather, it implies that in the analysis of the quantitatively identified associations between two variables analyzed

by, for example, correlation and chi-square statistics, the qualitative data will be applied to further explore the relation – that is, the introduction of further data on qualitative dimension of the phenomenon, allows the researcher to tap into and explain the generative mechanisms behind the quantitatively identified pattern (Clark, Creswell, Green, and Shope, 2010: 369; see also the discussion of statistical methods in chapter X).

To some extent, then, the approach applied here resembles the pragmatic stand on mixed methods advocated by, among others, David Morgan (2007: 71; Greene and Hall, 2010: 131; see also Danermark, Ekström, Jakobsen, and Karlsson, 2002: 151). According to Morgan (2007: 68) the paradigmatic distinction between qualitative and quantitative methodologies should be substituted by a pragmatic approach focusing primarily on the epistemological production of scientific knowledge, a move which enables him to bypass, at least to some extent, the metaphysical distinctions on which the distinction between quantitative and qualitative methodologies is based. The production of scientific knowledge cannot, he argues, be reduced to a distinction between induction and deduction, insofar as the move from theory to empirical data and back again is dialectical, and as such more resembles the process of *abductive* reasoning, just as the distinction between subjective and objective might be replaced with *intersubjectivity*, insofar as the very process of producing scientific knowledge is a continuous negotiation between the researcher, her respondents, other researchers, and the theories applied (Morgan, 2007: 71, 72). Furthermore, the distinction between context dependency and generality should be replaced with *transferability*, that is, can the results be transferred to other settings (Morgan, 2007: 72). The pragmatic approach applied here, may from a critical realistic perspective be criticized, insofar as it might, at least to some extent, understate the influence of the metatheoretical assumption on ontology (Danermark, Ekström, Jakobsen, and Karlsson, 2002: 153; Maxwell and Mittapalli, 2010: 147, 153). While it may be correct to argue that the research questions are not theory-neutral, and as such already centered upon certain metatheoretical assumptions, it might also be argued that neither of these approaches in themselves produce an appropriate foundation of the choice of methods, insofar as the choice must not only reflect the problem statement and research questions, but also the metatheoretical assumptions inherent to the theories applied in the construction of the research object, as well as the assumptions

inherent to the philosophical foundation, see the figure in the introduction to this chapter (see also chapter I).

Nevertheless, from a realist perspective, the question might also be what warrants the move from the claims to the postulate. That is, the postulates or the results of the studies must be founded in a number of claims and the move from claims to postulate – the theories on which this move is centered – implies a distinction between the *implied theories* expressed in the data and the scientific theories. The move is, in other words, centered upon a warrant which is mutually anchored in these implied theories and scientific theories. Moreover, the three dimensional construction of empirical data also implies a distinction between identifiable patterns across respondents – breadth – while the second dimension, the cognitive structures enables a depth perspective. The *missing* connections between these two dimensions will be added through theory – abstraction (Maxwell and Mittapalli, 2010: 156, 159; see also chapter I).

The approach described, here, is different from the one most often applied in the scientific fields on fairness research. The most *common* methodological foundation is a mono-methodological focus on quantitative studies (see, for example, Bandiera, Barankay, and Rasul, 2005; Van den Bos, Wilke, and Lind, 1998). In a rather simplified manner, the research can be divided into two broad traditions, one studying choices within frames either through questionnaires or laboratory studies (see, for example, Kahneman, Knetsch, and Thaler, 1986; Turillo, Folger, Lavelle, Umphress, and Gee, 2002), and another tradition focusing on game theoretic laboratory research (see, for example, Gächter and Riedl, 2005; Fehr and Gächter, 2002). While the results produced by these studies are interesting, it might be argued that some of the complexity inherent to qualitative studies is lost, insofar as the statistical inferences are based in existing theory, and as such the interpretive frame applied by the respondent when he/she answered the questionnaire or participated in the experiment is, to some extent at least, left open and filled out through the use of theories. While normative structures can be identified as patterns across respondents activated by different frames, it might be added that this activation can be viewed as the activation of some cognitive normative frames based on certain beliefs. To capture these beliefs one needs to inquire into the relation between the artificially constructed decision frame and the individual's perception of

this. Therefore, qualitative research methods might enable this study by inquiring into the foundation of these – that is, the reasons the agent had to formulate a particular judgment.

The mixed methodological approach may, as such, enable one to surpass some of the limits inherent to mono-methodological studies (Creswell, 2003: 22). However, it should be noted that the quality of the results produced is more influenced by the applications of the respective methodologies than the methodology in itself. That is, while a mixed methodological approach might solve some problems, it might also just double the problems insofar as the methodologies must also be applied in an appropriate manner (Hammersley, 2013: 14). The applied methods still have some limitations inherent to them, just as how they are constructed matter – I return to these aspects below. Hence, a mixed methodological approach is, in itself, not as important as the appropriate and correct construction of the respective methodologies applied. It might add to the complexity of the understanding of certain phenomenon, but the quality of the study is more influenced by the other dimensions mentioned (Maxwell and Mittapalli, 2010: 158).

### **9.3 Asking Sensitive Questions with Difficult Answers**

The aim of this study is to collect information about the respondents' perceptions on fairness. Additionally, the aim is also to understand how fairness is related with other horizontal phenomena such as social emotions (e.g. envy), morality, and more generally social relations. More specifically, it is as an inquiry into the foundation of judgments on horizontal and vertical fairness, and how these interact with one another. As such, these themes are not necessarily present at the conscious level, insofar as they are related with ingrained perceptions, just an inquiring into these areas may also be perceived as threatening, insofar as it bring forth, and as such confronts the individual with her own perceptions related with the social nature and horizontal dynamics of pay (Bradburn, Sudman, and Wansink, 2004: 80). That is, it is related with questions on “what I want for myself” versus “what I want for others” – in that sense, it is related with morality and social emotions on, for example, selfishness and envy. These themes may be viewed as threatening to the respondents' self-perception (Lee, 1993: 75; Tourangeau

and Bradburn, 2010: 338); hence potential biases may be the desire to appear fair, the desire to be a good person etc. Furthermore, due to the deep-seated nature of these themes, the respondents may also find it hard to answer these and this might call forth within them a desire not to appear non-reflexive, the themes may thus be perceived as provoking, and as such it may call forth within the respondents a hostile attitude.

Inquiring into these sensitive areas may be eased by constructing the questions in a non-threatening manner. *Depersonalization* is one approach, which may be enabled through the use of vignettes – stories describing other peoples' actions, beliefs, values, etc. (Finch, 1987: 110; Lee, 1993: 79). The use of vignettes makes it possible for the respondent to take on a third-person perspective, implying that the respondents are not evaluating their own behavior, beliefs, or norms; but those of the characters in the small stories. Furthermore, the wording of the questions may also aid the inquiry, insofar as the use of colloquialisms and everyday language may make the questions appear less threatening (Bradburn, Sudman, and Wansink, 2004: 81). Just as computer-aided surveys may also make the questions appear less threatening, because that the respondents are not being judged by another person while responding; hence they may be more open and less reluctant to answer what appears to be threatening questions (Bradburn, Sudman, and Wansink, 2004: 80).

As I described in the previous section, this study will draw on both quantitative and qualitative methods. More specifically, the study is built on a computer-aided survey and group/personal-interviews. The survey and the interviews will be based upon vignettes. In the survey, the third-person perspective may be easier to maintain insofar as the respondents will not have to get into contact with the researcher or other respondents' answers. Although, the group-interviews are based vignettes, the judgments made by the respondents are in focus, hence the third-person perspective might be more problematic to maintain. However, maintaining a focus on the vignettes, the hypothetical situations, and actions might enable one to collect data. The problem might, nevertheless, be that the respondents pass different judgments than if exposed to a first-person perspective. While the first-person perspective may be preferred, it might offset the depersonalization effect of the vignettes, and as such move focus from the characters described in the vignettes to the respondents themselves. In that sense, it might be argued

that a tradeoff can be identified between these two perspectives. Moreover, it might be argued that a first-person perspective would entail too much speculation, while a third-person perspective allows a level of generality to be implied, that is, the judgments passed by the respondents in the interviews may not necessarily express how they would act in particular, rather it may express how they believe the actors *should* act. A third-person perspective may therefore allow the researcher to access the normative foundation, albeit perhaps only in a limited sense – just as the nature of the interview type, group/personal interview, may influence the judgments passed by the individual respondents, insofar as the respondents in the group-interview may influence one another. As I will argue below, the aim of this study is not to study the formation of group consensus in the sense that the interviews will primarily be applied to complement and enable a deeper understanding of the phenomenon discussed and the cognitive structures on which this is based.

#### **9.4 Vignettes**

A vignette is a small story composed of different factors or variables which enables the researcher to collect data on some of the factors which might influence judgments on certain issues (Rossi and Anderson, 1982: 19; Finch, 1987: 108; Artzmüller and Steiner, 2010: 128). In that sense, the use of a vignette study enables the researcher to study judgments as contingent on certain factors, as contextualized, and as such, the assumption is that these judgments are vulnerable to framing effects – hence, the respondents do not hold non-contingent judgments; rather these shifts as the frames shifts (Finch, 1987: 106). Applying this perspective, allows the researcher to gather data on the patterns of judgments across respondents (Rossi and Anderson, 1982: 21).

A complete vignette study, allows the researcher to get some insight into the nature of the relation between these factors and the judgments made by the respondents, insofar as it allows the researcher to assess this influence based on direction, significance, magnitude, and differences across respondents and within case differences based on individual differences, just as it enables factor reduction (Rossi and Anderson, 1982: 33, 59). The vignette study, in other words, is centered on the assumption that the respondents hold different preferences over these factors and that this structure can be

explicated and studied (Rossi and Anderson, 1982: 20). A vignette study can, then, be viewed as a complex approach to opinion research in the sense that it allows the researcher to push past simple ratings, and to study the interaction between different factors influencing the judgment or opinion held by the respondents on certain issues (Finch, 1987: 105-106).

The construction of these vignettes implies two important choices, on the one hand, the researcher must choose the variables she identifies as important with regard to the issue she wants to study, just as she, on the other hand, must choose the number of items on the rating scale (Rossi and Anderson, 1982: 35, 49; Jasso, 2006: 342). The latter issue is related with the quantitative approach, thus I will discuss it below. The former issue, however, is more general in nature and will be discussed here, and in the next chapter. The choice of factors may be based on previous research from different sub-fields allowing the researcher to study the interaction between these factors. That is, drawing on previous research allows the researcher to get an intuitive idea about which factors might influence the judgments on certain issues, just as it may allow the researcher to produce hypotheses about the possible effect of certain factors upon the judgment. Introducing factors from different sub-fields, allows the researcher to get a more thorough understanding of the different factors which may influence the judgment, just as it allows the researcher to draw patterns, and as such to get a more complex insight into the frame on which the judgment on the issue of interest is based. In addition, the vignettes are not complete in the sense that they are influenced by the interpretative frame applied by the respondent (Finch, 1987: 110, 112).

The benefits of applying this vignette approach are that it allows the researcher to understand judgments as conducted within frames, just as it allows the researcher to get some insight into the nature of these frames. The disadvantages of applying this approach may be that the vignettes are constructed stories, that is, fictive scenarios based on the researcher's understanding of which factors may influence the judgment on certain issues, hence, it limits the field of possible factors which may influence the judgment, just as these fictive stories may make the respondents act more rational than they would do had these been real settings, insofar as the small stories allow the respondent to distance herself from the matters judged and apply a third-person perspective. This

may, however, not influence the study in the sense that the primary aim is to study decision frames. From a mechanism approach, it may be argued that the preferential structure explicated is not as important as the link between the context and the judgment, that is, gaining insight into the preferential structure is not sufficient to understand why certain judgments are passed, inasmuch as it solely uncovers the *inventory* of understanding. Taking this into consideration, allows the researcher to push past this inventory approach and to focus on the cognitive relations between the frames and the judgments, that is, trying to explain why certain patterns exist based on the interaction between the frame and the judgment. These explanations are built upon theory, and as such, they cannot be viewed as more than proxies of the actual cognitive mappings applied by the respondents. Moreover, the quasi-experimental nature of vignette studies can, from a critical realist perspective, be criticized because it assumes a closed-system approach to reality – one in which factors can be isolated – however, one may introduce a distinction between causal description and causal explanation, while the aim of the quantitative is to produce causal descriptions, the consequences which emerge from deliberative variation; the aim of the qualitative study (and the integration of the two studies) is causal explanation, the explanation of why certain patterns occur (Shadish, Cook, and Campbell, 2002: 9). Additionally, these causal explanations should be perceived as emerging in the interaction between theory and the two empirical studies, hence the introduction of several factors may *mimic* the nature of an open system approach in the sense that several factors are drawn in to explain the occurring phenomenon – the horizontal and vertical nature of reality (see chapter I).

### **9.5 Mechanisms and Analytical Strategy**

In the previous sections, I have argued that the aim of the quantitative study is to identify patterns across respondents, whereas the aim of the qualitative study is to identify cognitive patterns. Moreover, these patterns must be explained and connected through the use of theory. In accordance with this, I will now turn to the analytical strategy on which the following chapter is centered.

The aim of the present study is not just to identify patterns, but also to explain the foundation of these and the implications of these. The explanatory aim can be viewed as

having as its aim to explain why the present pattern emerged – that is, the aim is to explain the generative mechanisms behind the observed pattern. The multimethodological perspective applied here may aid this explanatory process, insofar as it allows one to apply a more narrow focus on particular judgments and the foundations of these (Mingers, 2004). In that sense, the aim of the analysis is to unfold the explanatory contexts within which the current behavior is based (see chapter III). The analytical strategy applied here implies an abductive and a *retroductive* approach – a continuous reflexive process, moving from theory to observed patterns, and from observed patterns to theory, just as it also implies a move from past (generative mechanism) to present (observed behavior), and from present to past (Danermark, Ekström, Jakobsen, and Karlsson, 2002: 95, 96). Whilst, statistical analysis primarily is concerned with describing probabilistic patterns associated with the behavior of some underlying population, and as such concerned with describing present-future connections associated with making statistical inferences; interpretive analysis is primarily concerned with describing interpretive patterns, and as such it is fixated in the present framed in the respondent's frames of understanding. The interpretive analysis enables an in-depth understanding of the observed phenomenon in the sense that it does not take these patterns as given, but inquires into the foundation of these. The retroductive approach, then, adds a further dimension to this by arguing that these interpretive patterns or frames on which the observed behavior is based are triggered by something – that is, the retroductive approach aims at identifying the abstract or transfactual conditions which might be invoked to explain the occurring phenomenon (Danermark, Ekström, Jakobsen, and Karlsson, 2002: 97; see also the discussion on mechanisms in chapter I).

Adopting a mechanism approach is more than just explaining causal chains – it is explaining the process of transformation which occurs when the cause changes into an effect (see chapter I). Explaining by mechanisms is to search for *clues* explaining why a particular observed pattern emerged. As such, the aim is not to directly explain in which situations a particular causal pattern is triggered, rather, indirectly, the aim is to explain the process of transformation – that which lies between the cause and the effect. Interviews may be applied to explain this transformation in the sense that these might enable the researcher to access, at least at the shallow level; the cognitive frames invoked in particular situations. Adopting a mechanisms approach entails, in other words, that the

researcher starts the analysis with a description of the observed pattern (present) and the potential causes of this (past) – tries to identify clues on the process of transformation which occurred in order to enable this process. These clues can, of course, be various things associated with explaining the observed causal relation. To reduce the complexity, it has here been chosen to focus on the association between the vignettes and the judgments made by the respondents. More specifically, the focus is on the relation between compensatory differences and judgments on fairness, social norms, emotions, and relational influence.

The analysis occurs in five basic steps (inspired by Danermark, Ekström, Jakobsen, and Karlsson, 2002: 109-110): *Step-1*, here the focus is on describing the events – that is, the aim is to present the results of the two empirical studies; *Step-2*, here the focus is on dissolving the observed tendency into several separate entities which enables that the phenomenon can be analyzed – that is, in order to analyze the relation between compensation systems, fairness, and relational influence, one might distinguish between *vertical fairness* associated with distributive and procedural fairness and based on understanding the relation between the context and the judgment; and *horizontal fairness* associated with causal agency and the act of differentiation; *Step-3*, here the focus is on explaining these different entities by reflecting upon the relation between the observed pattern, the entities which this has been dissolved into, and the theories invoked to explain these phenomena; *Step-4*, here the focus is on explaining why these patterns emerged by drawing on the theories introduced under step-3 – that is, to explain, for example, a particular response pattern related with judgments on causal agency, one may first explain the phenomenon from a theoretical perspective and then apply this theoretical explanation to explain the tendency in the data set, and, then, discuss the limitations of this explanation – under which conditions/in which situations is it likely that this is invoked to inform the respondent's judgment *vis-à-vis* when is it not likely that this is invoked. Because several theories may explain the same phenomenon of interest, situated at different strata, one may compare and discuss the explanatory power of these theories; and finally, *Step-5*, here the focus is on *concretizing* the results of the analysis in the sense that the aim is to focus on the practical implications of the results and their limitations.

The analytical strategy, just described, displays the logic of a sequential process from step-1 to step-5, however, the actual analysis, performed in the next chapter, will not be separated into five sections; rather, the analytical strategy should be viewed as the analytical process which occurs as the analysis occurs (the processual nature of analysis), thus I will not indicate when the different steps are invoked. That is not, of course, to argue that the steps occurs simultaneously, because they do occur separately, but it would not, here at least, provide any information to separate the steps into different sections of the complete analysis.

### **9.6 The Qualitative Study**

The aim of the qualitative study is to study the cognitive frames on which the different judgments of interest are centered. Additionally, the purpose of the study is to apply the results as further elaborations on the results identified in the quantitative study. As already explained, the purpose of the qualitative study is to collect in-depth descriptions of the way different judgments may be formulated by the respondents. The quantitative study is based on a survey, while the qualitative study is based on group/personal interviews. The respondents in the quantitative study are students enrolled at the Graduate Diploma in Business Administration education at University of Southern Denmark in the summer of 2012; while the respondents in the qualitative study are employees in three different small to medium sized Danish organizations. Moreover, the qualitative study is based on fewer respondents than the quantitative study due to the purpose of the study. The aim, then, is not to study the differences between the respondents belonging to different organizations or the differences between the two different groups of respondents, students at SDU versus not students; rather the aim is solely to study these judgments of interest from a quantitative perspective as patterns across respondents, and qualitatively as cognitive patterns. One underlying assumption of this study is, in other words, that the two groups of respondents shares a common set of norms on which these judgments of interest are based. This assumption might, of course, understate differences between the two groups, just as it might understate within group differences, and a study with unlimited resources could maybe have captured these differences, if all the respondents participating in the quantitative study had also been interviewed (Lieber,

2009: 221). This would, though, imply that the respondents' identity should have been disclosed in order to match the respondents with their responses, just as it would have entailed that the respondents should have been willing to participate in a follow-up interview. Although, this might have improved the results, it would have entailed a great amount of resources, just as it might be problematic to compare answers on a survey with answers given in an interview, insofar as the nature of these two response settings is very different. Therefore, in accordance with these considerations, the present research design was chosen.

### **9.6.1 Interviews**

The aim of the interview-research is to collect information about the cognitive structures surrounding judgments of certain issues. Interviews can aid the researcher in the collection on these issues and in particular in-depth interviews, personal and group, based on open-ended questions.

The aim, here, is not as such to study the language (discursive practice) applied by the respondents nor intersubjectively existing discourses; rather the aim is to explore conceptual mapping, that is, the beliefs and principles applied by the respondent when he/she elaborates on certain judgments (Giorgi, 2009: 136). In that sense, the approach applied here follows the hermeneutic and phenomenological approach advocated by, among others, Steiner Kvale and Svend Brinkmann (2009: 29). Furthermore, the approach is also closely related with the interpretative approach developed by Martin Heidegger (1926/1962), Hans G. Gadamer (1975), and Amedeo Giorgi's approach to a phenomenological research program in psychology (2009). According to this approach, then, the respondent's judgment on certain issues is connected with the frame within which this judgment occurs, that is, the respondent's judgment represent his/her perception on this issue, but this perception is based on certain beliefs and principles held by the respondent, which may be viewed as *by-products* of the respondent's private experience and appropriated frame of interpretation, see the next section on epistemological issues (Heidegger, 1926/1966: 210, 222; Gadamer, 1975: 282; Giorgi, 2009: 90, 96). Philosophical hermeneutics and phenomenological has their departure in the experience of the agent – the empirical (focus on the phenomenological experience) and actual (fo-

cus on the unfolding event) reality, and as such their scope is different from the one advocated in chapter I of this thesis, because here the focus was on collecting data on the generative mechanisms, phenomena which cannot be deduced from, at least not directly, experience. Although, Kvale and Brinkmann's (2009) approach is inspired by the before mentioned currents in interpretive philosophy, one could argue that these are primarily activated in a particular context related with the questions asked – that is, questions regarding experience may, of course, have a natural phenomenological dimension, whereas conceptual questions may not, insofar as conceptual questions are more general in nature and do not, necessarily at least, entail a description which takes as its point of departure personal experience. A method, however, which is based on obtaining data from actual respondents, must collect data on the empirical and actual reality, because the real world of generative is only reflexively attainable in the sense that these mechanisms are nothing but, theoretical abstractions on reality. One cannot, thus, collect data, directly at least, on these mechanisms – in the chapter on the philosophical foundation I argued that one could solely collect data on the results of these – and in that sense, one is necessarily confined to collect data on the empirical and actual reality. The aim of the conceptual interview is not, however, to collect data on the phenomenon as it exist in itself (Giorgi, 2009: 93), but rather to collect data on the mental representation of the judgment of interest as an attitude towards a particular issue – hence, the question becomes one of the knowledge implied not personal experience.

The metatheoretical differences aside, the application of interviews may enable the researcher to collect data on the foundation of the social judgments. Based on considerations regarding the nature of data needed and the more general aim of the interview research, the researcher may take several dimensions into consideration (Kvale and Brinkmann, 2009: 105, 107). On the latter point, the researcher may consider the kind of interview needed based on the aim of the research, insofar as this will influence the very construction of the interview. That is, if the aim is to collect personal experience on certain issues, the researcher must design the questions in a manner which allows her to collect this kind of data, whereas if the aim to test different theories these must be transformed into assumptions which can guide the researcher in her development of the interview-guide. The third kind of interview, the conceptual interview can be viewed as the combination of the two, insofar as the researcher's questions may sim-

ultaneously be driven by existing theories, just as she must also be able to distance herself from these in order to follow the cognitive mapping applied by the respondent – in that sense, the data constructed in this kind of interview is simultaneously constructed by the researcher and respondent in the interview-conversation (Kvale and Brinkmann, 2009: 105, 151). The aim, here, is not to collect data on personal experience nor to test theoretical assumptions, but to tap into the cognitive structures applied by the respondent when producing judgments on certain issues. The researcher is here not impartial, insofar as the interview can be viewed as a collective construction mimicking the dynamics of a normal conversation. The problem is, of course, that even though it mimics the normal conversation, it is constructed and purposively designed to collect data, and hence, the researcher must try to stir the conversation in the direction she finds appropriate with regard to the data she needs to collect (Giorgi, 2009: 95). Furthermore, the nature of the conceptual interview is closely related with a confrontational dialogue, in the sense that the researcher is trying to identify the interpretive frames applied by the respondent and inquire into these by identifying the basic assumptions applied by the respondent (Kvale and Brinkmann, 2009: 105, 158).

The aim of the interview may, in that sense, guide the choice of interview-type, just as it must influence the choice of data, and following from this the design of the interview-guide. The matter of which kind of data is needed depends also upon the analysis which the researcher wants to conduct. If the purpose is solely to collect personal experiences, then, the data needed must, in some way, be related with the respondent's life either expressed as stories produced by recalling past events or more phenomenological in nature, aiming at making the respondent recall how she felt at the time of event (Kvale and Brinkmann, 2009: 153; Giorgi, 2009: 93). The aim may also be to collect data about the opinions on certain general matters, just as it may also be, as it is here, to collect data on how the respondent produce these opinions – that is, the foundation of judgment. This approach, in other words, aims at pushing past mere collection of personal statements. The data needed here may not, necessarily, be narrative in nature, insofar as the aim is not to analyze the connection between concepts, story-lines, phenomenological issues, just as the aim is not to analyze the language applied. On the one hand, the aim is more concrete in nature, insofar as the aim is to collect data on the assumptions on which these judgments are centered, while, on the other hand, the

aim is also less concrete in the sense that the respondent may not be conscious of these assumptions, and as such the interview pivots around explicating these assumptions by getting the respondent to unfold her assumptions through elaborating on the foundation of her judgments. The data may still be viewed as simultaneously constructed by the researcher and the respondent, although the focus is less on mimicking the dynamics of a normal conversation, just as personal experience or stories – analogies, may also be viewed as important data (Morgan, 1996: 140, 145). The respondent may, however, not have any direct experience with the situations she is asked to judge, thus the researcher must aim at collecting data on more general cognitive frames – that is, frames which represents what the respondent finds, for example, to be *right*, and inquire into why this is found to be right. In that sense, this can be characterized as a non-structured interview, insofar as the researcher must be able to follow the cognitive path applied by the respondent, although the researcher may apply theories to guide her questions. As I explained above, to enable the gathering of data on this issue, it has here been chosen to draw on vignettes. The vignettes are designed to manipulate the information on which the respondent bases her/his judgments, although this information, of course, is only a part of the foundation, insofar as the respondent will base her/his judgment on beliefs, principles, etc., which needs to be explicated.

Furthermore, the researcher also needs to take into consideration the kind of interview she finds to be appropriate – personal or group-interview. Here, again, the choice is also influenced by, on the one hand, the aim of the research, and, on the other hand, the data needed, insofar as the personal-interview allows the researcher and the respondent to co-produce the data, while if other respondents are present, the data may also be co-authored by them (Kvale and Brinkmann, 2009: 105, 107; Morgan, 1996: 138). That is, while the interview allows the researcher to study cognitive mapping applied by one respondent, the group-interview allows the researcher to study the social nature of judgment (Morgan, 1996: 139). Although, the aim here is not to study the social construction of judgment, the application of group-interviews may also be viewed as useful method to gather data on threatening issues; insofar the interview-situation may appear less intimate, thereby allowing the respondent to adopt a third-person approach to her judgment. In that sense, applying group-interviews may also be viewed as a useful method to depersonalize issues in the sense that the respondent may feel less

threatened by the non-intimate interview-situation. This approach is, of course, not unproblematic insofar as it might prevent the researcher from gathering data on certain personal issues, because the respondent may be reluctant to disclose how she *really* feels about certain issues, just as the interview-setting in itself may prevent the researcher from collecting data on how the individual respondent relate to some issues insofar as she might try to adapt her judgment to that of the group or that of some particular respondent (Morgan, 1996: 140). However, the aim here is not on how particular respondents relate to certain issues, rather the aim is to collect data on some general patterns, just as the general aim is primarily to collect data which can aid the interpretation of the pattern identified through the quantitative approach.

### 9.6.2 Epistemological Considerations

The phenomenon of interest, here, is human judgment on issues related with fairness and social relations – hence, the aim is to gain some insight into the cognitive foundation of these judgments, the values and beliefs. The epistemological foundation of such an inquiry is, on the one hand, the experience of the respondents, and, on the other hand, the ingrained norms on which these judgments are founded. In that sense, the epistemological foundation may be characterized as being related with *lay knowledge*, understood as the way non-professionals relates to particular issues. While issues related with personal experience rests on the private experience gained by the individual respondents, the postulated existence of shared normative frames postulate the existence of shared cognitive frames across individuals (Giorgi, 2009: 96; see also chapter I). The subjectivist perspective might, then, at least to some extent, overstate the subjective nature of reality, insofar as it is a *social* reality and normative frames have extensions not just in shared pattern of behavior, but also in the particular ways that the interview respondents relate to particular issues (Danermark, Ekström, Jakobsen, and Karlsson, 2002: 164). And insofar as the aim of this study is to tap into the cognitive foundation of these issues, private experience may be considered as less relevant due to the fact that social agents have opinions on matters which cannot be reduced to personal experience. Tapping into these issues – matters which cannot be reduced to personal experience – requires the activation of shared cognitive frames. The epistemological foundation applied by the respondents when answering the interview questions cannot,

in other words, be reduced to private experience inasmuch as it implies that they transcend their private experience to frame the question in some general patterns, because the questions requires that they take on a third-party perspective.

In accordance with this perspective, a judgment may be viewed as an argument centered upon a distinct set of assumptions (beliefs and values) and claims. The aim of the interview is to tap into this foundation on which the judgment is founded – get the respondent to elaborate on her judgment in order to unfold the cognitive frame. To activate these general or shared normative frames, the vignettes will be applied (Finch, 1987: 110). However, the epistemological foundation cannot be reduced to this insofar as the argument unfolds itself in a social context – the interview-situation – implying that it is a negotiated epistemological foundation which emerges in the interaction between the individual respondent and the vignette, the individual respondent and the other respondents present, and in the relation between the interviewer and the respondents. Hence, the epistemological foundation cannot be reduced to personal experience due to the nature of the questions and the response-setting.

### **9.6.3 Interview-Guide**

The aim of the interview-guide is to generate the link between the research-questions, the questions on which the interview study is centered, and the interview-questions (Kvale and Brinkmann, 2009: 132). The study is based on gathering data about horizontal issues related with difference and differentiation, and how these issues are related with judgments on fairness. The aim of the study is to collect data on the cognitive structures on which these judgments on fairness are centered. The primary point of departure is the vignettes which will be discussed in the next chapter.

In accordance with this, the aim of this study is to collect data on the cognitive structures applied by the respondents when passing different judgments, that is, to collect data about the beliefs and principles on which these are based. Using the vignettes as points of departure, the respondents are first asked to judge whether they find the after-situation distributively fair, and secondly whether they judge the foundation of this to be fair – a judgment on procedural fairness. Thirdly, the respondents are asked to judge whether the receiving individual by doing some action which eventually leads to

the after-situation violates some horizontal norm. Fourthly and fifthly, the respondents are asked to judge the emotional significance of the reward to the receiving individual, and the emotional value of the compensatory difference to the non-receiving individual. And finally, sixthly, the respondents are asked to judge how this might influence the relationship between the receiving and non-receiving individuals.

Based on these judgments the respondents are asked to elaborate on the judgment they passed, that is, the aim of these questions is to collect data on the frames of interpretation applied by the respondents, the information they draw from the vignettes on which they base their judgment. Furthermore, by using why-questions the respondent is asked to justify her judgment, as such the aim is to collect data on the reasons behind these judgments. The aim is, in other words, to study the relation between the information provided in the vignette, the judgment, and the activation of certain frames of interpretation which, although they temporally precedes the judgment as a matter of sense-making, may only be accessed through the process of intersubjective justification. This should, however, not be understood as arguing that this allows the researcher to access the complete frame - rather the data must be perceived as fragmented and based on the respondent's perceived reason – what he/she believes to be the reason. That is not to argue that this is not the reason, just that the reason may be more complex in the sense that these patterns of justification primarily expresses the conscious dimension of the judgment – the surface level. While in-depth interviews may enable the one to push past this level, this might be perceived by the respondents' as threatening insofar as it might be perceived as intimidating or inquisitorial, hence the aim of the interviews is solely to collect data on this surface-level. Moreover, insofar as the aim is not to study the psycho-social construction of judgment, but solely cognitive patterns, moving past this level would not produce the knowledge needed here.

The following figure elaborates on the relation between the research question and the actual interview questions:

Figure 36

<b>Research question</b>	<b>Central question of this study</b>	<b>Main areas</b>	<b>Topics</b>	<b>Interview questions</b>
Which horizontal	Which horizontal	Direct causation:	Distributive fair-	Do you find the

problems can be identified in relation to individuality and individuality, and how is related with fairness issues?	problems can be identified in relation to vertical difference and horizontal differentiation, and how is this related with judgments on fairness?	is the after-situation produced judged as vertically fair?	ness	compensatory difference alright? Why/why not?
			Procedural fairness	Do you find the background of the compensatory difference alright? Why/why not?
		Indirect causation: is the after-situation produced by horizontal norm violating actions?	Horizontal social norms	Do you find that [name of the receiving agent] did something wrong by [the action done by receiving agent which indirectly brings about the after-situation]? Why/why not?
		Emotions: How important is the reward to the receiving individual, and how important is the compensatory reward to the non-receiving individual?	Emotional significance of receiving the reward.	How important do you think the reward is to [name of the receiving agent]? Why?
			Emotional significance of the compensatory difference.	How important do you think the pay difference is to [name of the non-receiving agent]? Why?
Relational influence: How may the compensatory difference influence the relationship between the receiving and non-receiving individual?	Relational influence.	How do you think the pay difference will influence the relationship between [name of the receiving agent] and [name of the non-receiving agent]? Why?		

Prior to the actual interview, the respondents will be asked to present themselves to the group and to the researcher – that is, state their name, tell about their job, their age, and what else they might find relevant. The aim of this presentation is, firstly, to make the respondents feel secure and confident, to get them relax, just as the aim, secondly, is also to enable later identification when processing the interviews. These data, then, are not a part of the data, but primarily have a practical function.

The interviews are structured in the following manner: firstly, the interviewer will present the research, his name etc.; secondly, the interviewees are asked to present themselves; thirdly, the researcher will hand out the vignettes and read the stories aloud, just as the interviewees are given some time to read the stories if they want; fourthly, the actual interview starts; sixthly, the interviewer asks if the respondents have anything to add or if they have any questions. In addition, all the interviews was conducted at the respondents' workplace in a lunch room or a conference room to secure peaceful, quiet, and familiar surroundings, just as the respondents along with the researcher was placed along a table to minimize the possible difference between interviewees and interviewer (Kvale and Brinkmann, 2009: 34).

#### **9.6.4 Respondent Group**

The group of respondents consists of 10 individuals, one group with 5 interviewees, two groups with 2 interviewees, and one personal interview. The respondents belong to three different organizations. 7 of the respondents belonged to organization Z, 2 of the respondents belonged to organization Y, and 1 respondent belonged to organization X. Furthermore, the group of respondents consists of 2 females and 8 males. The age ranges from mid-twenties to mid-sixties. 4 of the respondents held middle manager positions, while 6 of the respondents held non-management positions.<sup>39</sup>

Organization Z is a commercial-sector medium-sized organization which is part of a larger corporation. The core business of the organization is the production energy, fossil-based as well as green-energy. Organization Y is a medium-sized commercial-sector organization which is part of a large Scandinavian corporation which has its core business in retail to non-commercial consumers. Organization X, is a small-sized commercial-sector organization which is part of a larger corporation which has as its primary line of business, cleaning of waste water from industrial consumers.

The respondents were chosen based on the following criteria: they must have been working in the organization for some years or at least they must have experience with

---

<sup>39</sup> The interviews conducted at organization Z will not be used in the following empirical because they were conducted at an early state of the research process and based on a different set of questions. The file is, however, enclosed.

working in other organizations, this in order to secure that the respondents have some experience with working in an organizational context; the respondents must work in a manner which allows them to have daily or, at least, frequent interaction with their colleagues, this to secure that the respondents have experience with belonging not just to an organization, but also with being a part of a social community in a professional setting. The organizations were chosen based on the following criteria: they must be of some size which allows these professional horizontal relations to emerge, implying that they must be small to medium-sized in accordance with the EU definition stating that the organization must employ between  $50 \leq 250$  employees (EU, 2005: 14). Just as the organization may be part of a larger corporation, because this might increase the probability of it applying a compensation system similar to the one discussed in chapter II of this thesis. Furthermore, the organization must be located in Region Zealand.

Different sampling strategies may, of course, be applied depending on the purpose of the research and the data needed (Creswell, 2007: 128). In accordance with the above, the interviews are not based on collecting data on experience (as for example phenomenological studies are), the respondents of interest need not have any personal experience with the phenomenon, insofar as the aim is to collect data on some general phenomenon, the sampling strategy is, thus, based on the criteria described above (Creswell, 2007: 127). Further, it might also be argued that the application of the sampling criteria secures that the respondents are, at least to some extent, similar which might aid the group interviews, although some variation in the respondent may also contribute to the understanding of the phenomenon of interest.

The organizations selected in the first process was initially contacted by e-mail, see the enclosed mail. Approximately 100 organizations were contacted. This however, gave no results, thus 20 organizations were contacted by regular post, see the enclosed letter. This approach gave three results. To confirm the arrangement, a letter of confirmation was sent to the contact person in the organization. The respondents were then found in cooperation with the contact person in the respective organizations.

### **9.6.5 Transcription of Interviews**

The interviews will be transcribed in a meaning-condensed manner providing a summary of the respondents' discussions. This method has been chosen, because the interviews will not be applied in a linguistic analysis or some other analysis requiring full-transcription. The quotations which are applied in the analysis will be fully transcribed to allow a more thorough analysis of the cognitive structure applied by the respondent. The interviews was conducted in Danish, hence the quotations are translated into English. Translating from one language into another may, of course, be problematic insofar as some of the original meaning may be lost, however, because the focus here is not on the linguistic level, but rather on the general associative patterns – the important issue here is to maintain concepts applied by the respondents, while also trying to maintain possible connotations. Moreover, because the focus here is not on the linguistic level, conversational markers such as “mmh's” will not be transcribed (Kvale and Brinkmann, 2009: 181).

### **9.6.6 Quality Issues Related with the Qualitative Design**

The aim of this section is to discuss the problems of reliability and validity related with the qualitative study. Reliability concerns whether the study can be repeated and its results replicated. As will also be argued below, reliability is influenced by matters related with the respondent group and the research design. Reliability, as well as validity, explicates problematic aspects of qualitative research if applied in their *positivistic* meaning due to the distinct nature of qualitative research (see, for example, LeCompte and Goetz, 1982; Bryman and Bell, 2007: 410). This must, nevertheless, also be viewed in relation with the nature of the data actually collected in the study – that is, while qualitative data collected on some *local* phenomenon may cause some problematic aspects to emerge, the data collected here is based on exploring judgments on some rather general phenomena, hence the data collected here may not be as context specific or evanescent in nature as the data most often constructed using qualitative methods. This does not, of course, imply that the data collected here is general in nature and not vulnerable to contextual factors, although it might be argued that the interpretative frames invoked here are not, necessarily, specific to one organization, but rather they may reflect some more general cultural frames of interpretation.

The design of the interview-guide and the choice of interview-type may influence the reliability of the study in the sense that the application of open-ended questions and the spontaneity implied in the semi-unstructured approach may entail that the topics discussed and results collected are a result of the interaction between the interviewer and the interviewees, hence a different interviewer using a similar approach may not collect the same results (Kvale and Brinkmann, 2009: 172, 245). In that sense, the interaction between the interviewer and the interviewees may influence the results. Although, the interviews are rather spontaneous in nature, it should be emphasized that the interviews are designed to collect data on certain kinds of judgments implying that, even though, the researcher may not be able to completely determine, in advance, the questions she must ask to collect this data, it is still the same phenomena which are pursued across the interviews. Moreover, to capture the cognitive structuring of the respondents' judgment these were asked to elaborate on their judgments, just as they were asked to comment on (confirm) whether the conclusions drawn from these judgments reflected their understanding – that is, to secure intersubjective confirmation of the content of the judgment, the researcher tried to deduce, in a condensed manner, the conclusion which could be drawn from these and asked the respondent to comment on this understanding. This approach, then, is highly dependent upon the researcher's ability to capture these conclusions, just as it might be argued that the respondents when confronted with these conclusions are not, necessarily at least, conscious of the implications associated with these.

In regard to the transcription of the interviews, it may be noted that the translation of the interviews from Danish to English is problematic, insofar as some of the nuances might be lost. However, when statements are used in the analysis they are reproduced in text in a *complete* manner – meaning that they are reproduced in their dialogical form. The *original* Danish transcribed extracts from the interviews are enclosed, just as the interviews are enclosed in digital form.

The interviews are here solely applied to elaborate on the results from the quantitative study, hence the extracts from the interviews will be introduced, insofar as they are found to be relevant – that is, insofar as they contribute to the understanding of the phenomenon of interest. This implies, at least to some extent, that some of the nuances

which might be represented in the data set might be lost, because not all the respondents' opinions on the different issues are taken into consideration. To reduce this bias, the interviews were first transcribed in a meaning-condensed manner which summed up the main conclusions, and secondly relevant extracts was chosen and fully transcribed.

Furthermore, the quality of the study is also influenced by the choices explicated above – that is, is the method appropriate – does it enable the researcher to collect the needed data – are the questions relevant, and are the respondents relevant. The problem associated with collecting data on generative mechanisms is that these are only indirectly accessible in interviews. The questions asked must not be reduced to statements about experience, but must have a more general scope. Questions directed at gathering data on these general phenomena are, at least to some extent, problematic insofar as these are ingrained and surrounded by taken-for-granted knowledge, the method applied must, thus, explicate these cognitive structures. Moreover, because the aim is to collect data on these general phenomena, the respondents of interest should not be limited, insofar as it must be assumed that judgments on these matters are rather common in nature. In the above section some criteria on sampling was stated and it was stated that the respondents was identified in cooperation with the contact-person in each organization. However, in all the cases the respondents were chosen solely by the contact-person, based, hopefully, on the stated criteria and in some of the cases the contact-person chose himself/herself as a respondent. This is, of course, problematic in the sense that the researcher has no real influence on whom is selected and this might, as a result, bias the data. With regard to this research, it can be argued that because the phenomena of interests are general in nature and if the judgments are assumed to be common knowledge, the characteristics of the respondent group may not influence the validity of the study, although one should take into consideration that these characteristic may influence the knowledge collected about the subject matter, insofar as it may be assumed that conceptual differences exist across groups.

### **9.7 The Quantitative Study**

The aim of the quantitative study is to gather data on the patterns of judgments on fairness and relational influence. Unlike the qualitative study, the aim is here to collect data

across respondents. That is, the aim here is not to understand the nature of the particular judgment, rather the aim is to identify possible patterns across respondents which might be viewed as signifying the existence of shared frames of interpretation, and as such it may signify the existence of a shared normative set of expectations surrounding the interpretation of these particular situations described in the vignettes.

The quantitative study is based on a survey of the possible relation between fairness, horizontal social norms, and relational influence. The respondents in the survey were students, who at the time, was enrolled at HD at University of Southern Denmark.

### **9.7.1 Survey**

The present survey was based on an internet-based and self-administered questionnaire. The questionnaire was designed using SurveyXact. The survey was distributed to the respondent group by a collective e-mail send by the respective study secretaries attached to the different parts of HD, and the different study lines of HD-2. In the first e-mail, the potential respondents were asked to participate in the survey and a linked was provided at the bottom of the email which opened the survey. A second email was sent about three weeks later asking those who had not completed the survey to complete it. The two emails are enclosed, see appendix 3.

The choice of this particular survey method was based on two considerations, on the one hand, using a self-administered questionnaire allows the researcher to collect data on a rather large group of respondents in a time and cost efficient manner, just as it, on the other hand, can be viewed as a method which provides the respondents with the choice of whether or not they want to participate, but also with the choice of when/where they want to complete the survey – that is, it can be viewed as more convenient to the respondents, because they do not have to devote time in advance, but can complete it when they like, within the stated time line of course (Bryman and Bell, 2007: 241, 242). This convenience may, of course, also be associated with particular problems, insofar as the respondents cannot ask for help to interpret the questions (although email and phone-number was supplied in the survey) nor can the researcher probe and collect additional information which might aid her/him in the interpretation of the patterns, just as this particular method may also produce lower response rates and

missing data in the form of non-completed questionnaires (Bryman and Bell, 2007: 242, 243).

The choice of the survey methodology and the choice of the particular method – a self-administered questionnaire – are like the choice of interview form, associated with the aim of the research and the data needed. The data needed here, is patterns of judgments that can be identified across the respondents, just as the aim of the research is to study how the patterns of judgments across three different main areas, but also how these judgments might be interrelated. The particular approach, in other words, must enable the production of these data. This also entails that the design of the questionnaire must enable the production of this kind of data – that is, from a pragmatic perspective, there must be a relation between the data produced and the data needed. Moreover, the aim of the research must also be related with the research question that the researcher wants to answer based on the results of the data produced by the particular method. The research question addressed here is the same as the one addressed in the qualitative study, that is, which horizontal problems can be identified in relation to individuality and individualization or vertically introduced difference and horizontal differentiation. In contrast, however, to the qualitative study which had as its aim to collect data on individual cognitive frames, the aim is here, at a higher aggregated level, to collect data on patterns – that is, the results of these cognitive frame, the actual patterns of judgments across respondents.

These considerations are not just related with the choice of methodology and method; they are also related with the very design of the questionnaire – that is, the choice of questions, the formulation of these questions, and the scales of measurement. The latter choice will be discussed in the next section. As already explained, the point of departure for both the qualitative and quantitative study was the three vignettes, presented in the next chapter, just as the questions related with each vignette was centered upon collecting data on judgments associated with fairness, horizontal norms, emotions, and relational influence. Because the aim of this survey was to collect data on patterns of judgment, closed questions were applied without the “don’t know” possibility. This was done in order to make the respondent *guess* even if he/she might not hold any opinion on the matter. The choice of closed question was done in order to enable later trans-

formation of these individual judgments into patterns, although open questions might have provided a richer or nuanced understanding of the subject it might also have complicated the process of identifying these patterns, insofar as these different judgments would have to be encoded in a particular manner and because each respondent might express their judgment differently, coding bias might follow, just as open questions might be viewed by the respondent as more time consuming, and as such it might lower the response rate (Krosnick and Presser, 2010: 267; Bryman and Bell, 2007: 259). Furthermore, the use of closed questions may also be appropriate due to the specific nature of the respondent group – students – because students are more used to guessing, hence the questionnaire must encourage the respondent to guess, rather than just moving on to the next questions (Krosnick and Presser, 2010: 268). This seems to be particularly relevant in studies of more general phenomena in which no right or wrong answer exists, though the respondent may not hold any opinion on the particular matter he/she might still be able to pass a judgment on these matters (Krosnick and Presser, 2010: 284). In other words, the respondent may not hold any pre-produced opinion on the matter or he/she may not be certain how they should judge, thus they should be encouraged to use their intuition and simply judge what they think is the right answer. In the present study, the respondents were encouraged to just follow their intuition, because no right/wrong answers exist on these general matters, see the enclosed questionnaire.

In the questionnaire, the vignette was posted at the top of page and followed by the questions to reduce navigation problems associated with going back and forth between different pages, just as it might be easier to respond, if the information needed is on the same page. The questions were grouped into four categories: (1) the two first questions was related with judgments on fairness; (2) the third question was related with horizontal norms; (3) the fourth and fifth question was related with the emotions of the receiver and the non-receiver; and (4) the sixth question was related with judgments on relational influence. The grouping of questions into these four categories was done intentionally to ease the respondents' ability to recall the information provided in the vignettes, insofar as the first category draws on one type of judgment associated with similar information, that is, the categories are based on likeness and relatedness of judgments, just as they are centered upon judgments based on similar information in order to reduce possible recall errors and to reduce the task difficulty. To reduce the task diffi-

culty, the questions across the three vignettes are ranked in the similar order, just as they are semantically similar in wording with only few differences (Krosnick and Presser, 2010: 291, 293). This may, however, also be problematic insofar as latent interaction effects may be built into the order of the questions, that is, a particular pattern may emerge across the three vignettes due to some association between the questions, just as similar wordings of the questions may make it harder for the respondents to cognitively distinguish between the three different vignettes – the respondents may simply continue the pattern which they found appropriate in the first vignette (Tourangeau and Bradburn, 2010: 335). To prevent this, the sequence of the questions could have been changed under each vignette. Although, this might have reduced a potential bias it might also have enhanced the task difficulty, insofar as the respondents would have to apply different cognitive processing schemes at the three vignettes – because the order of the questions differed they would have to focus on different parts of the vignette to retrieve the information needed, whereas by standardizing the order of questions enables the respondents to learn which piece of information they need to answer the questions, and as such it might make the task appear easier (Krosnick and Presser, 2010: 292). In addition, applying the same semantic structure the questions, across the three vignettes, may also produce fatigue insofar as the respondents may find the lack of variation demotivating.

Besides these questions on different kinds of judgments related with the information provided in the three vignettes, the respondents were also asked to answer some general questions about themselves, their line of work, etc. In the questionnaire, these questions followed after the questions related with the three vignettes. This was done intentionally to secure that the respondents, at least, completed the first part of the questionnaire – that is, due to concerns about fatigue and the requirements on reading and processing of the information provided in the vignettes, it was found relevant to divide the questionnaire into two parts with different levels of importance. The first part was found most important; hence it was given priority over the second part. Because the vignettes may be viewed as more challenging it might also be problematic to put them first, insofar as they might discourage some respondents because they may find them too hard or too time consuming – while had the order been reversed the respondents would first have been faced with some easier questions, although this might have led to

the consequence that fewer respondents would have completed the questions related with the vignettes due to fatigue after having answered several questions they would have been required to read a text and to process the information provided in it (Krosnick and Presser, 2010: 292).

### **9.7.2 Measuring-Scale**

The choice of measuring scale – the number of items on the scale – must reflect the cognitive mapping of the phenomenon (Krosnick and Presser, 2010: 268). That is, the choice of measuring scale must match the one normally associated with the phenomenon, insofar as the judgment made by the respondents is the result of a transformation of information provided into a judgment which, then, must be transformed into, for example, a number on a rating scale (Krosnick and Presser, 2010: 269; Tourangeau and Bradburn, 2010: 337). Furthermore, the scale must appear ordinal in the sense that it is a continuum which progress from one end-point to another (Kuzon, Urbanek, and McCabe, 1996: 265; Jamieson, 2004: 2017).

The choice of the number of items on the rating scale is also associated with validity and reliability inasmuch as it is related with the accuracy of the measure and whether the measure measures what it is supposed to (Krosnick and Presser, 2010: 272, 273). A measuring scale with more than two items is more accurate, because it allows nuances to be expressed, just as it may be more valid because it captures differences in judgments which may reflect true opinions more accurately. However, the problem is to identify the natural cut-off point – the point at which adding an additional item produces no further information. A scale which provides the respondents with three options may, for example, be more accurate, although a neutral point may just as well reflect “don’t know”, hence it may generate ambiguous information. A scale, on the other hand, which provides the respondents with more options, while also allowing a neutral point might be more accurate – to secure a natural neutral point, equidistant end-points might be applied (Krosnick and Presser, 2010: 274). This implies that an odd number of items may be preferred. While 5 and 7 points scale may provide too little space at each part of the neutral point, an 11-points scale may be too wide, thus here a 9-points scale will be applied. Although, 5 on a 9-points scale may be viewed as a neutral point, only the end-

points can be labelled insofar as the qualitative differences may be problematic to capture semantically (Krosnick and Presser, 2010: 275). The lack of semantic coding may be problematic, insofar as it might introduce interpretation errors, because the respondents may map the qualitative differences differently than the researcher (Kuzon, Urbanchek, McCabe, 1996: 266; Jamieson, 2004: 2017). To reduce the possible problems associated with this, the 9-points scales will be collapsed into a relevant fewer number of items in the analysis. Collapsing the items into fewer, but broader categories may, of course, reduce the nuances associated with a larger number of items, but it might make the interpretation easier, just as it might also enable an analysis insofar as this also depends on the distribution of answers across the categories.

### 9.7.3 Epistemological Considerations

The epistemological considerations related with a questionnaire concerns the base of responses or judgments – that is, the epistemological foundation of these judgments. To understand the epistemological foundation of survey responses, one might initially focus on the cognitive process of answering survey questions (Tourangeau and Bradburn, 2010: 318). According to Tourangeau and Bradburn (2010: 317), one may distinguish between four different levels which may not, necessarily, occur in a sequential form; rather the process may go back and forth between the different levels. The first level is *comprehension*; this concerns the respondents' process of interpreting the question, that is, the process of reading the question and cognitively combine the string of words into a coherent sentence expressing a question, and the process of interpreting the researcher's intended question (Tourangeau and Bradburn, 2010: 319). The second level is *retrieval*; this concerns the respondents' process of recalling the information needed to answer the question, that is, it concerns the respondents' ability to recall, for example, a particular event and the emotions associated with this particular event (Tourangeau and Bradburn, 2010: 321). The third level is *judgment*; this concerns the respondents' ability to pass a judgment, that is, it concerns the respondents' process of passing a judgment (Tourangeau and Bradburn, 2010: 322). Some questions may be easier to answer, because the respondent has prior experience with these questions, for example, age or gender, whereas other questions may draw on, for example, pre-existing opinions on general matters (Tourangeau and Bradburn, 2010: 323). Although,

the information needed to answer the question may be retrieved from memory, the actual judgment is constructed implying that it is the result of a process of reasoning, that is, the answer is based on the information retrieved from memory, just as it is based on information triggered by the context (the previous questions), and general or socially appropriate values ascribed to the particular issue on which the respondent is asked to pass a judgment (Tourangeau and Bradburn, 2010: 323). The fourth level is *reporting of the answer*, this concerns the respondents transformation of his/her judgment into, for example, a particular number on a measuring scale, as discussed in the previous section (Tourangeau and Bradburn, 2010: 324).

From what has been argued so far, it seems that the epistemological foundation of surveys is a combination of recalled information; the context of information provided by the questions and the vignette; and the influence of pre-existing frames of interpretation which may not only influence the process of judgment, but also the reporting of the answers insofar as the respondent may edit her/his response. In that sense, like it was the case with interviews, the epistemological foundation of the data collected through a survey is based the interaction between different epistemological sources, just as the actual judgment is the result of a process of reasoning. The answer, in other words, must be viewed as the result of this cognitive process preceding it, hence the patterns identified across the respondents can be viewed as generated by the cognitive process of each respondent implying that the identification of patterns necessitates the existence of shared frames of interpretation across the different respondents. That is, the patterns identified represent not only similar interpretations of the information provided in the vignettes, but also similarity in patterns of reasoning which might be viewed as an indication of shared norms guiding their process of judgment formation.

#### **9.7.4 Respondent Group**

As the table below shows the data set consists of 49.74 percent men and 50 percent females. The majority, 61.3 percent, of the respondents are between 19 and 33 years of age, while 32.5 percent is between 34 and 48 years of age. 23.3 percent of the respondents have an educational level corresponding to public school, a youth education or a General Upper Secondary School Leaving Certificate, while 63.1 percent have an

educational level corresponding to a Short or a Medium Higher education, and 13.6 percent have a Higher education. Additionally, 66 percent of the respondents have 1 to 11 years of experience, 26.2 percent have 12 to 23 years of experience, and 7.9 percent have more than 24 years of experience.

As explained above, all the respondents were enrolled at HD at University of Southern Denmark. 29.5 percent of the respondents were enrolled at the first part of the education – the Graduate Certificate Program (HD 1) and 70.5 percent of the respondents were enrolled at the second part – the Graduate Diploma Program (HD 2). Students from HD 2 are, then, overrepresented in the data set. 46.7 percent of the respondents were enrolled at, or if HD 1 would like to attend, a HD 2 with emphasis on economics, e.g. Management Accounting or Accounting, while 53.3 percent were enrolled at, or if HD 1 would like to attend, a HD 2 with emphasis on non-economics, e.g. Marketing Management or Organization and Leadership. At the time of the survey, about 400 students were enrolled at HD 1, 200 at HD 2 with economic emphasis, and 275 with non-economic emphasis. The total population then is 875 students with a response rate on  $(343/875)$  39.2 percent.

Furthermore, 98.4 percent of the respondents were at the time of completing the survey employed. 90.9 percent of the respondents were employed in the private sector, while 9.1 percent in the public sector. The majority, 82.4 percent, of the respondents had a contracted work time equal to 37 hours per week. In addition, 59.9 percent of the respondents described their primary work activities as being characterized by administration or accounting – e.g. in-house consulting, HR related, financial controlling etc. – 27.4 percent of the respondents described their primary activities as management, while 12.9 percent described their primary work activities as corresponding to support and front line activities, e.g. IT or marketing. Lastly, 59 percent of the respondents describe the compensation type attached to their job as salary without some variable component, e.g. bonus, 32.1 percent had a compensation type with salary and some kind of variable component, e.g. bonus or provision, while 9 percent had a compensation type described by either hourly wages or piece rates.

Figure 37

		Count	Percent
<b>Gender</b>	<i>Man</i>	95	49,74
	<i>Female</i>	96	50
<b>Age</b>	<i>19-33</i>	117	61,3
	<i>34-48</i>	62	32,5
	<i>49+</i>	12	6,3
<b>Level of education</b>	<i>Public School and Youth Education</i>	41	23,3
	<i>Short and Medium Higher Education</i>	111	63,1
	<i>Higher Education</i>	24	13,6
<b>Years of professional experience</b>	<i>1-11</i>	126	66,0
	<i>12-23</i>	50	26,2
	<i>24+</i>	15	7,9
<b>HD PART</b>	<i>HD Part 1</i>	56	29,5
	<i>HD Part 2</i>	134	70,5
<b>HD Direction</b>	<i>Economic</i>	85	46,7
	<i>Non-economic</i>	97	53,3
<b>Management responsibility</b>	<i>Yes</i>	51	27,6
	<i>No</i>	134	72,4
<b>Agreed work time, hours per week</b>	<i>30-36,75</i>	10	5,3
	<i>37-37,5</i>	154	82,4
	<i>38+</i>	23	12,3
<b>Reward type</b>	<i>Salary with a variable component</i>	46	59,0
	<i>Salary without a variable component</i>	25	32,1
	<i>Wage</i>	7	9,0
<b>Job characteristics</b>	<i>Administration and Accounting</i>	111	59,7
	<i>Management</i>	51	27,4
	<i>Support and frontline functions</i>	24	12,9
<b>Sectors of employment</b>	<i>Private sector - Commercial</i>	165	88,7
	<i>Private sector - Non-profit</i>	4	2,2
	<i>Public sector - State</i>	13	7,0
	<i>Public sector - Regions</i>	1	0,5
	<i>Public sector - Municipalities</i>	3	1,6

### **9.7.5 Quality Issues Related with the Quantitative Design**

In this section, I will discuss the potential issues related with reliability and validity of the quantitative study. Reliability concerns, on the one hand, whether the results can be repeated over time, just as it, on the other hand, concerns whether applying a similar methodology leads to the same results – that is, are the results consistent over time, and are the measures accurate (Golafshani, 2003: 598; Bryman and Bell, 2007: 163). In that sense, reliability is partly concerned with whether the results are consistent over time, and are partly concerned with whether the results can be repeated as a matter of design features, for example, the design of measuring scales, and the preparing of the data for analysis. If the results are consistent over time it would demand that the respondent group did not change their opinion on the matters, while this may seem unlikely (Golafshani, 2003: 599), due to the general nature of the judgments these may be more stable over time. In addition to this, it might also be argued that the group of respondents is rather special; hence the results may not be replicable. In regard to the design of the survey, it may be argued that the use of a 9-points measuring scales can be viewed as problematic, insofar as the respondents may interpret it differently, although as will be discussed in the next chapter, the pattern identified across respondents within each vignette case is rather consistent which might be viewed as an indication of shared interpretation. Furthermore, the consistency of judgments across the three vignette cases may be a result of the design of the questionnaire in the sense that to reduce perceived task complexity, the questions were formulated in a similar manner, just as the order of the questions was the same across all three cases. As argued above, this potential error follows as a trade-off between standardization which allows learning or non-standardization which might produce more reliable measures. The choices made in relation to the preparation of the data for analysis will be discussed in the next chapter.

Validity concerns whether or not the measures applied actually measure what they should measure and whether the choice of methodology is in accordance with the data needed (Golafshani, 2003: 599; Bryman and Bell, 2007: 165). The first question concerns whether a survey based on vignettes allows one to collect data on patterns of judgment, while the latter question concerns whether the data produced is in accordance with the data which should be produced. A survey-based vignette study is fairly commonly applied to collect data on normative issues in social science. The application

of this methodology enables one to collect data on patterns across respondents, however, the data obtained reflect not actual judgments; rather they reflect third-person judgments on these issues of interest implying that they may not be viewed as statements of how the respondent might feel/act should he/she find himself/herself in a similar situation as the ones described in the vignettes. Furthermore, due to the cognitive psychological processes of the answering process, it might also be argued that the respondent before reporting his/her judgment will edit it to fit her/his self-image or perceived normative expectations. While first-person judgments might be preferred, these can also be viewed as problematic insofar as they would imply that the respondent should try to identify with one of the characters in the vignette, and as such it would rely on the respondents' ability emphasize – to put themselves into the position of one of the characters and to activate the emotional foundation which might be problematic if the respondent have no personal experience. Moreover, insofar as the aim is to collect data on normative expectations it might also be argued that these are not, necessarily, contingent upon personal experience; rather they must be identified as patterns of shared interpretations across a similar group of respondents. Thus, from the perspective described here, the data collected seems to be in accordance with the data needed.

### **9.8 Methodological Delimitations**

As I explained in the above sections, the aim of this study is to collect data on the cognitive structuring of judgments as well as patterns of judgments across respondents. Furthermore, the study is based on an integration of interviews and a survey. The results of the survey will generate information on the patterns of judgments across the respondents, whereas the interviews will generate information on the cognitive structuring on which these judgments are based. The primary aim of the analysis is to analyze the generative mechanisms based on an integration of the results from the individual studies and the theory.

The focus of the qualitative study is based on vignettes and has as its aim to collect data on a general phenomenon, judgments on outcome fairness. The introduction of vignettes allows the respondents to take on a third-person perspective, and as such it may be one approach to collect data on the general cognitive structures which goes into

the formation of judgments on fairness, just as the approach may allow one to collect data on rationality of judgment, insofar as the respondents takes on this third-person perspective. This approach, then, does not allow one to collect data on phenomenological issues related with the cognitive structuring of these judgments, for example, the implications of emotions and neither does it allows one to collect data on the influence of other kinds of experience related with judgment as based on or influenced by subjective factors. Moreover, because the focus in the interviews was on the vignettes, data on the context was not collected, insofar as the thick-description here is not based on the relation between the context and the object of interest; rather, the thick-description is an abstraction on the possible mechanisms which may have brought about the judgment of interest. As explained above, this is not to argue that contextualization is not interesting, all that is claimed here is that the particular may be of less importance when the analytical and philosophical perspective is changed from social constructivism to critical realism.

If a phenomenological-hermeneutical approach had been applied, the data of interest would have been rich descriptions of the context and the respondents, insofar as the phenomenon could have identified as generated in a dialectical process between the individuals' frames of interpretation and the contextual frame. In addition, the collecting of this data would have changed the focus of the analysis from the general existence of the phenomenon to a focus on the subjective and social-psychological bases of judgment. Had this approach been applied, the interview-type applied should also have been changed from the confrontational-conceptual interview to the in-depth interview either in the form of personal or group interviews. This approach could also have been based on, for example, a grounded theory perspective as the philosophical perspective, while this approach might have identified other sources of information through its qualitative-inductive approach it would have implied that more interviews should have been conducted, just as other kinds of qualitative data should have been collected, for example, internally-distributed documents such as the organizations' compensation policies, personnel-handbooks, etc. However, this kind of data is difficult to gain access to due to the sensitivity of the topic, hence the perspective could have contributed positively to the analysis, but pragmatic limitations in regard to data access might have influenced the overall quality of the study.

The focus of the quantitative study was to collect data on the patterns of associated with the main areas; vertical fairness, horizontal fairness associated with differentiation, emotions related with receiving and not receiving, and relational influence. As I will argue in the next chapter, the vignette set is too small to identify the possible of the different independent variables – the contextual factors which might have influence on the respondents' judgments. Had a larger set of the vignettes been tested, the analysis could have focused on determining the relation between the descriptions provided in the vignettes and the judgments across respondents, just as it could have focused on how different factors influences the judgments made by the individual respondent applying a multivariate data analysis. This would, at least to some extent, have influenced the results generated in the analysis positively; however, on the one hand, the number of vignettes might have discouraged some respondents from participating implying that the number of potential respondents should have been increased to secure an acceptable response rate; and, on the other hand, if the quantitative study had been applied without collecting any qualitative data it would have been problematic to combine with the philosophical foundation of this thesis, at least to some extent, insofar as the quasi-experimental nature of the method presupposes a set of assumptions on the nature of reality which might not be fully supported by a critical realist perspective as argued above. In addition, instead of the integrated approach applied here, the qualitative study (or the quantitative study) could also have been applied to generate a set of hypotheses which, then, could have been tested in another study. This would also, at least to some extent, have contributed positively to the results, although the results produced by this approach would have been less rich in nature, insofar as collecting data on the same questions using qualitative and quantitative data allows one to study the phenomenon from two different dimensions – as patterns across respondents and as cognitive structures related with the passing of the judgment. This data richness would have been lost had the above mentioned approach been applied.

## Chapter X

### Exogenous differences, Endogenous Differentiation, Causal Agency, and Relational Influence

#### 10.1 Introduction

In the previous chapter, I discussed the methodological foundations of the present chapter; hence I will now turn to the analysis and discussion of the empirical results. In the previous chapters, I have been discussing the theoretical foundation of the normative expectations inherent to the horizontal and vertical relations and how these might influence the perceived fairness of the organization's compensation system. In the present chapter, I will focus on the relation between fairness framed as relational status quo, fairness framed as past/present status quo, causality, and relational influence. In that sense, the aim is to study three of the components in the foundation of outcome claims – the relational, temporal, and cognitive normative foundation. As such, the primary aim is not test the theories discussed in previous chapters; rather, the primary aim is solely to illustrate empirically some of the theoretical points and, if possible, to identify some limits to these. Of course, the data set is too small to allow for the construction of anything but illustrations – illustrations which might, however, provide some basic understanding of how the phenomena, which have so far, in this thesis, only been theoretically constructed, might be cognitively structured.

#### 10.2 Presentation of the cases

As I discussed in the previous chapter, the point of departure for the empirical analysis was three small stories/vignettes framing three particular situations based on which the respondents were asked to make different judgments. The three stories are showed below.

Case-1: The *Peter and Søren case*.

Peter and Søren are both employed as project-consultants in the same organization. Søren has been employed in the organization for five years, while Peter has only been employed for one year. They both have a cand.merc. Peter's monthly salary is 34,064 Dkr., while Søren's monthly salary is 49,064 Dkr. They are both viewed as being competent to their
---

job, just as they both supply an effort which is above the expected level. In addition, they share an office. In the previous period Peter was, after hard work, so lucky as to become part of a strategically important project. Søren also tried to become a part of this project, but did not succeed. The project turned out to be a success and each member of the group received a one-off bonus on 15,000 Dkr. This entailed that Peter's salary last period was 49,064 Dkr.

Case-2: The *Pernille and Jeanette case*.

Jeanette and Pernille are both employed in the same organization. Jeanette was hired 4 years ago, while Pernille was hired 2 years later. Both employees hold a HD, just like they are both viewed as experts in their field. Jeanette's average monthly salary is 39,550 Dkr., while Pernille's average monthly salary is 36,550 Dkr. Their job includes that the two must work together on the preparation of the department's budgets and other administrative tasks. Pernille have in the current period attended an on-the-job training course on a new IT-system and this has qualified her to be coordinator when the system is to be implemented next fall. Pernille choose to accept the company's offer because she perceived as a career opportunity. In relation to her attendance of the course, Pernille has received a monthly qualification allowance on 9,137 Dkr., this entails that her monthly salary now is 45,687 Dkr. Jeanette was also offered to attend the course but refused to do so.

Case-3: The *Anders and Jørgen case*.

Jørgen and Anders are both employed as engineers in the same organization. They were both hired 5 years ago to aid the startup of a new department. Since their hiring, both Jørgen and Anders have worked on creating a well-functioning department. Both Jørgen and Anders are recognized as supplying a satisfactory performance. Jørgen is perceived as an expert in his field and very theoretically gifted, thus Anders often come to him for advice. Anders, on the other hand, is very pragmatic and has primarily been concerned with the management tasks. At the last performance appraisal, Jørgen was highly recognized for his theoretical knowhow. Anders was also positively recognized for his management skills and he was offered a take a MBA (Master in Business Administration – a relative expensive education).

### 10.3 Variables

In the present section, I will discuss the choice of the independent and dependent variables. The independent variables are those which generates the frame – the composition of the small stories – whereas the dependent variables are the different judgments made by the respondents. Furthermore, all the small stories describe a before-situation and an after-situation.

### 10.3.1 Independent Variables

Each story is composed of similar sets of independent variables which generate the frame. These stories are not complete, as already discussed in previous chapter, entailing that the *complete* story is made in the interaction between the respondents' understanding of the stories and the matters they, in a popular wording, read into these stories. As discussed in the previous chapter, the aim of the interviews was to get an understanding of these matters which the respondents might read into the stories – that is, when they fill out the interpretative gaps – or make sense of the information provided.

Basically, in accordance with the theories discussed in chapter V and VII, a situation is relationally structured on difference between an initial distribution and a new distribution which, temporally, can be associated with a before situation and an after situation. In accordance with equity theory, the relational structuring is designed in a manner which takes into account rank ordering – hence, horizontally, an outcome difference is defensible when associated with input factors such as the ones I will discuss below. Furthermore, in chapter VII, I also identified that fairness claims may be cognitively assessed based on a comparison between before and after, and that the past binds the future, thus if the after situation is to be judged as fair it must maintain the relational proportionality which might not necessary be reduced to matters related with rank ordering, insofar as it might also be based on outcome expressed either in absolute terms or qualitatively transformed into a relational distinction between: above/below/equal. This latter point, in other words, can be viewed as an attempt to combine the Fehr-Schmidt model on relational fairness with the Kahneman-Knetsch-Thaler model on the temporal dimension of fairness.

The four independent variables: *educational background*, *job title*, *competence level*, and *seniority* all relating to the relational structure between the parties and, as such, they are all related with the dimension of relational comparability and with the proportional understanding of fairness. From the theories on equity discussed in chapter V, it seems to follow that relational symmetry on these dimensions should lead to compensatory equality, while relational asymmetry should lead to non-equal compensation. These independent variables can, then, take on three dimensions: above, below, or equal. However, because I only study three cases, it is impossible to capture the effect of these, insofar as one would have to study all the dimensions and all the different

combinations of these; hence these variables should primarily be viewed in combination with each other, as generating the relational structure. In addition to these variables, which in accordance with equity should generate the normative foundation of entitlement, the theories discussed in chapter VII, predicts that a past/present status quo may also influence the normative foundation of perceived entitlement, thus to capture this effect, *outcome or salary* can be applied. The outcomes of both parties in the before-situation and the after-situation are explicitly stated in each story. The important matter here is not, as discussed in VII, whether the employee is entitled to a particular difference in absolute or percentage terms; rather the important matter here is, whether the employee is entitled to a particular relational qualitative difference expressed as above, equal, or below. The five independent variables, discussed above, all relates to the matter of difference, however, as discussed in chapters II, VI, and VIII, differentiation might also matter, therefore to capture this effect, a sixth variable is introduced, *intentionality*. Intentionality captures the effect that the after-situation may either be *causally* related with the actions of the receiving individual and, as such, the agent is indirectly causing the after-situation implying that the cause is endogenous to the horizontal relation, whereas the after-situation may also be brought about by purely exogenous matters as in case-3. Of course, intentionality can never be entirely pure as described in chapter VI, insofar as causal agency presupposes control and the employee differentiating herself from the other employees can never be blamed for bringing the after-situation about, because she cannot control whether her actions are recognized, see also chapter II. The employee who indirectly brings the after-situation about may; however, be violating some other social norm as also described in the two examples provided in chapter II, see also chapter VIII.

### **10.3.1 Dependent variables**

In the present subsection, I turn to the dependent variables. The matters on which the respondents were asked to judge. The *first* matter on which the respondents were asked to judge was related with the distributive fairness of the after-situation – that is, to what extent the respondents judged the after-situation to be fair or unfair. In the survey, a qualitative nature was introduced with the use of a Likert-scale, while in the interviews the respondents were first asked to judge the fairness of the after-situation, and

then to elaborate on the foundation of this judgment, see chapter IX for a further discussion. *Secondly*, the respondents were also asked to judge, to what extent they found the foundation of the compensatory difference which, directly, brought about the after-situation to be fair or unfair. The respondents were, in other words, asked to judge the procedural fairness of the exogenous cause of the after-situation. *Thirdly*, the respondents were asked to judge to what extent the receiving individual (he/she who benefitted from the compensatory difference) violated some normative expectation inherent to the horizontal relation between the receiving and non-receiving individuals by indirectly bringing the situation about. *Fourthly* and *fifthly*, the respondents were asked to judge the emotional significance of the compensatory difference to the receiving and non-receiving individual, respectively. From what was discussed in chapter V, envy and guilt might be viewed as drivers of the emotional experience of receiving and not receiving, just as the emotional experience might also be influenced by the hedonic experience (pain/pleasure) of receiving and not receiving relative to the cognitive expectations held by the individuals, see chapter VII. However, to reduce the complexity of the judgment, the respondents were only asked to judge whether they believed it to represent a positive or a negative significance to the receiving and the non-receiving individual, respectively. And finally, *sixthly*, the respondents were asked to judge to what extent they believed the after-situation would influence the relation between the receiving and the non-receiving individual. Here, one should expect the drivers to be either “fairness for me” and “fairness for other” or as Fehr and Schmidt (1999) argued, envy and guilt (see chapter V). Furthermore, it could also be argued that the emotional experience is related with whether or not the after-situation was brought about as a result of a norm-violating action, see chapter VIII.

#### **10.4 Two Basic Scenarios**

In order to fully comprehend the complexity of exogenously produced *differences* and endogenous *differentiation*, I will distinguish between two basic scenarios, one related with relational status quo, and another related with indirect causal agency. Again it must be emphasized that due to the size of the data set any comparison made cannot be viewed as anything but an illustration.

#### **10.4.1 Exogenous Differences**

Here it might be questioned whether or not relational status matter – that is, is there a difference between changing from a before-situation characterized by compensatory inequality to one characterized by compensatory equality, as exemplified in case-1; from one characterized by equality to one characterized by inequality – exemplified in case-3; or from one characterized one type of inequality to another characterized by a different form of inequality. The theories discussed in the chapters V and VII are primarily focused on the comparability and the temporal dimension of entitlements, and as such these theories focus primarily on differences in comparable input or past/present relational structure. However, Fehr and Schmidt (1999) only discussed equality as relational structure, while Gächter and Riedl (2005) primarily focused on how past compensatory differences may bind present negotiations. None of these studies, then, focused on whether the nature of the relational change makes a difference.

#### **10.4.2 Endogenous Differentiation**

In chapter II, it was argued that self-differentiation might be viewed as antisocial behavior, and in chapter VIII this was further discussed in relation with the human basic need to belong. The question might, however, be whether perceived intentionality actually matters – as Lysgaard (1961/2010) seems to argue. In other words, does it make a difference whether the indirect cause of the after-situation is perceived as a norm-violation or not? Case-1 compared with case-2 allows a study of whether the nature of the act of different matter – that is, is there a difference between the indirect cause being participation in a project-group or seizing a career-enhancing opportunity. Comparing case-3 to either of the other two cases allows one to study the difference between intentionality and non-intentionality, insofar as the receiving individual in case-3 does not intentionally try to bring about the after-situation. The aim is here, then, primarily to understand whether indirect causal agency matter, but also whether the nature of it matters in terms of intentionality/non-intentionality, and kind. This seems particularly important in regard to the new compensation theories discussed in chapter II, just as it seems to be a gap in the theories on equity discussed in chapter V.

## 10.5 Analysis and Results

The aim of this section is to present the results of the analysis. The analysis will be structured in the following manner: firstly, the data will be analyzed drawing on descriptive statistics and interview quotations; secondly, the possible relations between dependent variables will be analyzed.

Before turning to the actual analysis, some methodological comments must be made with regard to the methods of analysis applied. In this chapter, I will primarily draw on odds ratio analysis. The odds ratio will, here, be applied in two different ways with two different aims. As a general concept, odds ratio describes the likelihood of one event occurring compared to another. Odds ratio calculated in relation to frequency distributions, describes the odds of one event occurring compared to some base-line event, here the event that the respondents found the situation neutral. Odds ratio, on the other hand, calculated in relation to cross-tabulations, describes the likelihood of some event occurring given the respondents' judgment on a different matter (Knoke, Bohrnstedt, and Mee, 2002: 160; Liao, 2004: 760), for example, it may be applied to describe the likelihood of judging that a pay difference will have no or little relational influence, while also judging the outcome difference to be distributive fair, relative to finding the pay difference to have a negative relational influence and finding the outcome difference distributive unfair. In that sense, the odds ratio, here, describes the association between the dependent and the independent variable, whereas the chi-square test primarily allows one to determine whether a relation exist or not, that is, it allows one to reject the null hypothesis stating that no relation exists between the two variables in the population (Knoke, Bohrnstedt, and Mee, 2002: 142; Bryman and Cramer, 2011: 203), and as such it does not provide any information on the nature of the association between the two variables. Furthermore, because the chi-square test is based on expected observations and the number of observations in each category (Bryman and Cramer, 2011:207), the test of significance will be based on Fisher's exact test if the expected count of observations is below 5 percent (Knoke, Bohrnstedt, and Mee, 2002: 104). In addition to this, the chi-square test is applied as a non-parametric test analysis of relation due to the nature of the data – because these are ordinal – which also entails that when correlation analysis is applied, this will also be non-parametric and based on *Spearman's rho* (Bryman and Bell, 2003: 364).

Odds ratio is not based on hypothesis test, instead confidence intervals are calculated as proxies of statistical significance, insofar as large confidence intervals represents a lower level of precision, whereas small confidence intervals represents higher levels of precision, just as a confidence must not overlap the null-hypothesis (odds ratio = 1) stating that no relation exists (Knoke, Bohrnstedt, and Mee, 2002: 160-161). Moreover, to support the statistical significance of odds ratio, the p-value may be estimated based on the z-value (Liao, 2004: 761). The odds ratio is defined within the following range:

Odds ratio = 1: The independent variable does not affect the odds of outcome

Odds ratio > 1: The independent variable is associated with higher odds of outcome

Odds ratio < 1: The Independent variable is associated with lower odds of outcome

Intuitively odds above 1 are easier to interpret; hence odds below 1 can also be expressed as the inverse value,  $1/\text{Odds ratio}$ . Odds ratio > 1, can be interpreted as stating that the judgment expressed in the independent variable increases the odds of judging in a particular way – the outcome of the dependent variable. Odds ratio < 1, on the other hand, can be interpreted as stating that finding something to represent a particular judgment on the independent variable decreases the likelihood of judging in a particular way – the outcome of the dependent variable. Moreover, odds > 1 may also be interpreted as expressing that the likelihood of one event occurring compared to the baseline is higher, whereas odds < 1 may be interpreted as expressing that the likelihood of the event occurring compared to the baseline is lower – or equivalently, the event is either more likely or less likely to occur compared with the baseline.

### **10.5.1 Exploring the Cases**

The aim of this subsection is to describe the trends in the data set through descriptive statistics and interview quotations. Consider the table below:

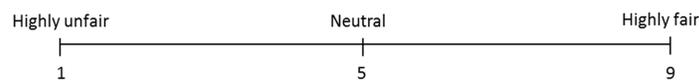
Figure 38

**Distributive Justice: "Please evaluate whether you find the compensatory difference fair?"**

	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative percent</i>	<i>Odds ratio with middle category as base-line</i>	<i>90% Confidence intervals (Lower bound, Upper bound)</i>
Unfair	8	4	4	0.3	(0.17, 0.66)
Neutral	23	11	14	1.0***	(0.61, 1.63)
Fair	185	86	100	49.7***	(34.51, 71.58)
<b>Case-1</b> Total	216	100			
Unfair	6	3	3	0.2	(0.08, 0.37)
Neutral	29	15	18	1.0***	(0.65, 1.54)
Fair	162	82	100	25.8***	(18.51, 16.00)
<b>Case-2</b> Total	197	100			
Unfair	4	2	2	0.2	(0.07, 0.46)
Neutral	19	10	12	1.0	(0.59, 1.71)
Fair	167	88	100	66.0***	(44.26, 98.41)
<b>Case-3</b> Total	190	100			
Peter and Søren c:	127				
Missing data					
Pernille and Jeane	146				
Anders and Jørgen	153				

In the table, the nine items in the Likert-scale has been collapsed into three broader categories. Although, some of descriptive quality of the scale is lost, the distribution of answers and the size of respondent group make it necessary to reduce the number of items. The category “Unfair” covers the items from “Highly unfair” 1 to 3; the category “Neutral” covers the items 4 to 6; and the category “Fair” covers the items 7 to 9 “Highly fair”. The grouping is centered on the following interpretation of the Likert-scale:

Figure 39



To the right of the category “highly unfair” and to the left of the category “neutral”, the judgment is either that it is less unfair or mildly fair, in the sense that as the phenomenon becomes less unfair it also becomes more fair, just as when it becomes less fair it also becomes more unfair. Because of this interpretative ambiguity, I have here chosen to reduce the qualitative differences associated with differentiating qualitatively between higher/lesser degrees of fairness and unfairness. Moreover, this interpretation also seems to be in accordance with the distribution in the data set, in the sense that the majority of the respondents found the outcome distribution to be 7-9, thus these items must have been interpreted as the categories cognitively associated with fairness, while the

less frequently used items 1-3 and 4-6 must have been associated either with unfairness or, at least, some category which does not capture the degree of fairness.

A similar pattern can be identified across the three cases, insofar as in case-1, 86 percent of the respondents found the outcome difference to be distributive fair, in case-2, 82 percent of the respondents found it fair, and in case-3, 88 percent found it fair. In accordance with this pattern, in case-1 the odds of finding the outcome difference to be fair compared to finding it neutral are 49.7; in case-2 the odds of finding it fair to neutral are 25.8; and in case-3 the odds are 66.0.<sup>40</sup> Across all cases, then, the likelihood of finding the outcome difference fair to neutral is significantly higher. The likelihood is highest in case-3 and lowest in case-2. In case-3, the receiving individual was offered the possibility of attending a MBA-education, while in case-2 the receiving individual obtained a lasting quality bonus. The respondents were solely asked to judge to what extent they found the after-situation fair/unfair – implicit in this judgment may, of course, be a comparison between before and after; however, no data was collected on how the respondents judged the before-situation, so the possible implications of such a possible implicit judgment cannot be identified - from a relational fairness perspective. In case-2, the receiving individual went from obtaining a lower outcome than the non-receiving individual and in case-3; the receiving and the non-receiving individual initially received an equal outcome and in the after-situation the receiving individual obtained a higher outcome than the non-receiving individual. From a pure temporal status quo perspective, none of the cases should be fair; insofar as they all entails that the before is unequal to the after-situation, just as from a pure relational status quo perspective, the outcome difference in the before-situation is not maintained in the after-situation. In terms of horizontal proportionality, in case-2 the receiving individual increased the quality of her input by attending an on-the-job course, while in case-3 the receiving individual differentiated himself because he focused on management related tasks, and in case-1 the receiving individual differentiated his input by taking part in a project team.

---

<sup>40</sup> The odds in this section are calculated based on the following procedure: The probability of event A, B, C, Odds =  $([A/1-A]/[B/1-B])$ ; Confidence interval: (lower bound =  $\text{LN}(\text{Odds}) - Z\text{-coefficient} * \text{SE}$ ), (lower bound =  $\text{LN}(\text{Odds}) + Z\text{-coefficient} * \text{SE}$ ) [reported in antilog];  $\text{SE} = \sqrt{([1/B \text{ (marginal)}] + [1/A \text{ (marginal)}])}$ . The P-value is approximated using the Z-score = Odds/SE

Because the actual difference in the distribution of responses across the cases is so small, it cannot be statistically determined whether the differences in odds are statistically significant, thus it might solely be speculated that reward differences following from the receiving agent's active agency are more likely to be found fair than reward differences emerging from the receiving agent *just* accepting an opportunity offered to her. The lack of response differentiation across the three cases, however, seems to point to an interesting point, namely that equality is not preferred to equity and neither is the opposite present, implying that the respondents in this survey did hold any preferences with regard to this distinction – at least none which could be transformed into differences in the response distribution.

In accordance with the analytical strategy described in the previous chapter, the central question is here: why is the compensatory difference across all the cases judged to be distributive fair – that is, how may this pattern be explained or phrased differently, which mechanism enables the transformation of the information provided in the vignettes change into judgments on distributive fairness? To get a sense of this, consider the two interview quotations:

*Case-1: The Peter and Søren case:*<sup>41</sup>

*Interviewer:* Do you find the pay difference to be okay?

*Interviewee:* Yes, I think it is. He did do something extra.

*Interviewer:* What do you mean by “something extra”?

*Interviewee:* Well, I just mean that he did something besides his normal work.

*Case-2: The “Pernille and Jeanette” case:*<sup>42</sup>

*Interviewer:* Do you find the pay difference to be okay?

*Interviewee:* I do.

*Interviewer:* Why?

*Interviewee:* She participated in that in-job-training course. That is not something one does for fun. You do it because you expect to be rewarded.

Based on these two quotations, it seems that the respondents, at least the respondents in the interview, focused on the information that Peter and Pernille both actively did some-

---

<sup>41</sup> Filename: Organization Y, time code: 00:01:41

<sup>42</sup> Filename: Organization Y, time code: 00:08:24

thing to bring about the pay difference. This seems to be in accordance with the equity theory, in the sense that relevant differences should trigger different compensatory outcomes. However, based on the two quotations it may also be argued that the distributive fairness judgment was not based upon a direct comparison between the inputs and outcomes of the receiving and non-receiving individuals; rather the judgment seems to be based upon the information about what the receiving agent did, and in particular that he or she did something besides their work – they did *something extra*. In accordance with what was argued earlier, both agents showed active agency. On the one hand, this *something extra* may be viewed as a difference in relation to the receiving agent's *normal* work, just as; on the other hand, this *something extra* might also be understood as relational differentiation, insofar as the receiving agent did something which differentiated her/him from the non-receiving agent. From this it seems to follow that positive deviations in terms of either the receiving agent's normal work or compared with some other might influence whether a compensatory difference is found fair, at least, from a third-person perspective. That is, the extra, the unrequired, or that which surpasses expectations is considered as a legitimate foundation of distributive fairness judgments. Whether this is the result of *spontaneity* (as in case-1) or simply *seizing an opportunity offered* (as in case-2) may be of less importance, inasmuch as both cases implies agency in the form of choice – the choice to do something extra – and this choice is, as argued by the respondents in the interviews, accompanied by an expectation about being rewarded. Hence, another interpretation might also be that it is considered to be fair if it is aligned with the expectations accompanying the actions of the receiving agent, see chapter II. In accordance with this, then, it seems to follow that, at least, in cases in which active agency is involved, the more or less static factors influencing the fairness of the outcome claim, as discussed in the previous chapters, may be less important. Hence, one might speculate whether agency can explain why the fairness expectations related with maintaining the temporal and relational status quos seems not be present here in neither the interview or in the survey responses. If this is the case, the dynamic characteristic associated with agency offsets fairness as emerging from temporal and relational comparisons. Following from this is also the assumption that if this is indeed the case, fairness in a behavioral perspective may be more important than comparisons, at least, from

a third-person vertical perspective. The horizontal dimension will be discussed below.

Figure 40

**Procedural Fairness: "Please evaluate whether you find the foundation of compensatory difference fair?"**

	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative percent</i>	<i>Odds ratio with middle category as base-line</i>	<i>90% Confidence intervals (Lower bound, Upper bound)</i>
Unfair	10	5	5	0.3	(0.14, 0.46)
Neutral	36	17	22	1.0***	(0.68, 1.48)
Fair	167	78	100	17.3***	(12.78, 23.44)
<b>Case-1</b> Total	213	100			
Unfair	6	3	3	0.1	(0.07, 0.29)
Neutral	35	18	21	1.0***	(0.67, 1.48)
Fair	156	79	100	17.1***	(12.59, 23.33)
<b>Case-2</b> Total	197	100			
Unfair	5	3	3	0.2	(0.07, 0.36)
Neutral	30	16	18	1.0	(0.65, 1.53)
Fair	157	82	100	23.9***	(17.22, 33.23)
<b>Case-3</b> Total	192	100			
Peter and Søren	130				
Missing data					
Pernille and Jeanette	146				
Anders and Jørgen	151				

Turning now to procedural justice, a similar tendency, as identified with regard to distributive fairness, can be identified. As the percentage distribution of the responses seems to argue, the majority of the respondents in all the cases found the foundation of the compensatory difference to be procedurally fair. The odds of finding the foundation of the compensatory difference fair to finding it neutral in case-1 are 17.3, 17.1 in case-2, and the odds in case-3 are 23.9. The likelihood of finding the compensatory difference procedurally fair to finding it neutral is, in other words, statistically significant.

The slight difference in the distribution of responses between the distributive fairness variable and the procedural fairness variable may be explained with reference to the nature of the judgment, insofar as it might be more problematic to judge whether the foundation is fair or not. This might also be viewed in relation to the descriptions provided in the vignettes which do not elaborate on the foundation, but solely state what it is, for example, in case-1 the foundation was Peter's participation in the project. In that sense, to judge whether the foundation is fair or not may demand more information than to judge whether something is distributive fair. An information substitution may, in other words, occur implying that information related with different kinds of fairness judg-

ments may be applied to make sense of the fairness of some phenomenon (Van den Bos, 2001: 68). One of the respondents argues on this matter in relation to case-1:

*Interviewer:* Do you find the foundation of the pay difference to be alright?

*Interviewee:* What do you mean?

*Interviewer:* The decision on which it is based.

*Interviewee:* Do you mean because he participated in that project?

*Interviewer:* Yes for example.

*Interviewee:* Yah, that is okay.

*Interviewer:* Why?

*Interviewee:* That is a tricky question. But I believe that it has something to do with knowing the background, I mean you have a sense of why.

*Interviewer:* So, it is important to know why?

*Interviewee:* Yes, perhaps it is not that important in relation to one's own pay. But when it comes to others'...then one would like to know why and whether it is based on other matters.

*Interviewer:* Other matters?

*Interviewee:* Yah, things which might not be okay – things unfit to be discussed.

*Interviewer:* What could that be for example?

*Interviewee:* Well, if someone was good friends with the boss.

*Interviewer:* So, knowing why have an influence on whether you find it to be okay?

*Interviewee:* Yes, because then you have the opportunity to decide for yourself whether you find it to be okay. That is hard if you do not know why. Then, your imagination takes over.

*Interviewer:* So, it is easier to accept if you know why?

*Interviewee:* Yes, but, of course, you must also find the background okay, it is not enough to know why.

*Interviewer:* So, if you know the background and find it to be okay it is easier to accept?

*Interviewee:* I would think so. But you might, of course, still be disappointed and angry. Then you need some days to come to reason with it.

*Interviewer:* Towards whom do you direct this disappointment and anger?

*Interviewee:* Not at anyone in particular. Well, of course, the ones who you blame, those who took the decision.

*Interviewer:* But not your colleagues?

*Interviewee:* No, not directly at least. But you might discuss it with others or your family. But it passes rather quickly because you have so many other things to care about.”<sup>43</sup>

From what is argued here it seems that one can distinguish between two important features of information about the foundation of a compensatory difference; on the one hand, the effect of knowing why, and, on the other hand, the effect of accepting the reasons. That is, knowing why is important, but the agent still has to accept the reasons on which this is based. Finding something to be procedurally fair, in other words, is a combination of these two effects, knowing why and accepting the reasons. As also argued by Lawler (2003), knowing why limits the possibilities of producing beliefs based solely on imagination or non-factual causes, and as argued by Folger and Martin (1986) it limits the possibility of counterfactual belief formation, see chapters II and VI. However, it may not limit the agent’s emotional response, albeit the effect of these emotions may be limited and the duration short. The effect of these emotions quickly wears off, just as the agent comes to accept the decision if she finds the reasons on which it is based to be fair. From what is argued here, it also seems to follow that information about and acceptance of reasons may be of particular importance when it comes to judging compensatory differences, insofar as it provides information about why the other obtained her outcome, and as such, acceptance of these reasons might be viewed as acceptance of the difference.

This seems to be aligned with previous findings by, among others, Van den Bos, Lind, Vermundt, and Wilke (1997: 1035, 1036, 1038, 1041; 1998) and Van den Bos (2001: 66, 67) who found that procedural fairness matters in situations when little or no relational information is held, insofar as it serves as a heuristic on which the employee may base her outcome perception – that is, from a vertical perspective, fairness judgments may be perceived as foundational of the employees’ perceptions about how they are treated, just as the fairness of past interaction may influence present fairness judg-

---

<sup>43</sup> Filename: Organization Y, time code: 00:10:13

ments. In the cases presented here, the respondents had information about the input and outcome of the receiver and non-receiver, just as they had information about what directly caused the pay difference to be brought about. However, in addition to this, it might be argued that procedural fairness may also be applied as a horizontal measure of fairness, insofar as knowing why some pay difference exists is also important, that is, social comparison may provide some of the foundation, but this foundation needs, it seems, to be supported by further information on why the difference exists in the first place. Knowing why may indeed have a *fair procedure effect*, insofar as knowing why and finding this to be fair might influence how one evaluates one's own outcome, in the sense that a fair treatment might be interpreted as signals of respect and inclusion, hence the employee might reciprocate this by accepting less than optimal outcomes (Van den Bos, Lind, Vermundt, and Wilke, 1997: 1035; Van den Bos, 2001: 68; see also chapters III and VI).

Knowing why may also be important because a compensatory difference is directly brought about by some decision preceding it and, as such, it is not the result of some random event or something beyond the control of management; rather it represents a decision based on some reasons, and it is those reasons which come to generate the foundation of whether or not one is treated fairly. This can also be viewed in relation to what was argued in chapter VI on the difference direct and indirect causality and its relation with responsibility, just as it might also, in a vertical perspective, be associated with Fehr and Falk (2002) theory on the implications of quasi-moral expectations such as reciprocity, see chapter IV. In the Fehr-Falk model as well as in Rabin's (1992), the intention was via a second-order belief transformed from a cognitive belief to a socio-emotional representation of the other and her intentions behind addressing another agent in a particular manner. It should be noted that in difference to the relational perspective on organizational fairness, based on the distinction inclusion/exclusion, the models based on reciprocity seems to be centered upon a different distinction: good/bad and respect of governing social norms/transgression of governing social norm, thus the symbolic message carried in the behavior of the parties in the vertical relation, its manifestation, and decoding is significantly different from one another, see chapters III and IV. What is expressed here, in the respondent's representation seems, however, to be more factual in nature and related with knowing the facts behind the decision. Gaining

access to this information may, of course, be problematic and especially in cases when the organization might have no explicit or only a very general compensation policy or in cases when the evaluators' behavior deviates from the organizations compensation policy. That is to say, the agents' access to information about their own and others process as well as outcome may be limited. Nevertheless, knowing why may be important due to the fact that compensatory differences are brought about by direct causal agency of somebody and, as such, the decision may be expected to be based on some justifiable criteria. From what is argued here, employees may be more rational than is sometimes assumed, insofar gaining factual information on the reason matters, just as the reasons must be defensible and not just representations of biased evaluations. Furthermore, although compensatory differences might stir up emotions, the employees are not, at least not according to the respondent, likely to act on them or if they do, they direct them towards those who are blameworthy – those who make the decisions.

Combining the results, it might be argued that concerns about maintaining a relational status quo does not influence fairness judgments, insofar as greater differences should be expected across the three cases in terms of fairness judgments. In that sense, from the perspective of the impartial spectator – a third-party – relational status quo does not seem to influence fairness judgments. It might, then, be speculated whether or not the distribution of answers across the categories would have been different had the respondents been asked to identify with the non-receiving part. However, as argued by Messick and Sentis (1979), fairness norms may hold prerogative over preferences, implying that fairness is a matter of governing fairness norm. Here, it seems that “participation in a project-team”, “pursuing a career-enhancing opportunity”, and “different areas of expertise” are found to be generating fair reasons for compensatory differences. This seems, in other words, to be in accordance with Lawler's (2003) argument on the dominance of rational acceptance, see chapter II. Against this it might be argued that Lawler (2003) is referring to acceptance from a first-person perspective – that is, *A* rationally accepts her own compensation as well as that of *B*. In Messick and Sentis' (1979) study as well as in Gächter and Riedl's (2005) study, the respondents were asked to take a first-order perspective and identify with one of the parties, therefore the results found here may also be due to the perspective, insofar as it might be easier to produce non-biased judgment from a third-person perspective. Nevertheless, it seems that the

results here contradicts the ones found by Kahneman, Knetsch, and Thaler (1986a; 1986b), although they did not directly study horizontal entitlement, it seems that past/present entitlements inherent to the horizontal dimension does not influence judgments on the fairness of compensatory differences.

Figure 41

**Norm Violation: "Please evaluate whether you find the receiving agent to have violated some unwritten rule?"**

	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative percent</i>	<i>Odds ratio with middle category as base-line</i>	<i>90% Confidence intervals (Lower bound, Upper bound)</i>
No norm violation	179	84	84	38.5***	(27.23, 54.43)
Neutral	26	12	96	1.0***	(0.63, 1.58)
Norm violation	9	4	100	0.3	(0.16, 0.58)
<b>Case-1</b> Total	214	100			
No norm violation	190	96	96	776.0***	(367.46, 1638.77)
Neutral	5	3	98	1.0	(0.35, 2.84)
Norm violation	3	2	100	0.7	(0.20, 2.20)
<b>Case-2</b> Total	198	100			
No norm violation	168	88	88	53.8***	(36.99, 78.18)
Neutral	22	12	99	1.0	(0.61, 1.64)
Norm violation	1	1	100	0.1	(0.01, 0.40)
<b>Case-3</b> Total	191	100			
Peter and Søren	129				
Missing data					
Pernille and Jeanette	145				
Anders and Jørgen	152				

From the fairness variables, I now turn to the judgments on whether or not the receiving agent violated some horizontal social norm. The items in the Likert-scale have, here, also been collapsed into three different categories. The items “No norm violation at all” 1 to 3 have been combined in one category, 4 to 6 have been collected in a second category, and 7 to 9 “High degree of norm violation” in a third category. Across the three cases, it can be seen that the majority of respondents found that no norm violation had occurred. However, in the case-2, the odds of finding that no norm violation had occurred to finding it neutral are 776, while in, for example, case-1 the odds are 38.5. Some of the difference might be due to the slightly higher percentage of respondents who judged it to belong to the “Neutral” category than belonging to the category “No norm violation” in case-1.

Another explanation of this difference might be the fact that in case-2, the non-receiving individual was offered the opportunity, but declined it; hence it might be speculated whether causal agency on the part of the non-receiving agent influences

judgments on norm violation. That is, causal agency on the part of the non-receiving agent may offset possible concerns about social norms governing horizontal relations. Although, it cannot be interpreted as stating that equal distribution of opportunities matter, it can be interpreted as stating that if the non-receiving agent self is to blame, then this offsets possible social norms. Social norms on, for example, relational equality may not govern situations in which the non-receiving agent brings the situation upon her through causal agency. In regard to what I argued in chapter VI, whereas the receiving agent by differentiating herself cannot be blamed for bringing some particular after-situation about – at most indirectly for producing a certain kind of actions, the non-receiving agent may in some cases be directly responsible, at least, insofar as a direct causal relationship can be established between her action and the after-situation. On this one of the respondents in the interviews argued:

*Interviewee:* She must have seen it coming. The other has, of course, nothing to do with it. But of course you still have to behave appropriately.

*Interviewer:* So, you believe that makes a difference?

*Interviewee:* She cannot blame anyone...only herself.

*Interviewer:* Does that make it easier?

*Interviewee:* No, on the contrary.

*Interviewer:* So it would have been easier to accept if it was introduced from the outside?

*Interviewee:* Perhaps...now, she can continue to blame herself.

*Interviewer:* So you believe it will influence their relationship negatively?

*Interviewee:* Not necessarily. But the other will be a constant reminder of what she did. But, of course, you do not just blame someone else.

*Interviewer:* Earlier you mentioned that one was not allowed to make the other one sad/upset. In this history, you say that the person only has herself to blame – does that make a difference?

*Interviewee:* You are still not allowed to make the other one sad.

*Interviewer:* So, under all circumstances you have to behave appropriately?

*Interviewee:* I would mean so.

*Interviewer:* Do you believe it will change her perception of whether her own pay is alright?

*Interviewee:* Not necessarily. But if you blame yourself ...it will probably be expressed in some way.

*Interviewer:* So you believe it might change her perception?

*Interviewee:* Perhaps. But not because it is wrong. Simply because she feels sorry for herself. ”<sup>44</sup>

Based on this it might be argued that causal agency on the side of the non-receiving may offset some relational norms, but not those related with basic social behavior such as not making anyone sad or upset. Furthermore, even though the non-receiving declined the opportunity she may blame herself – that is, she might regret her actions, though at the time she did not desire to accept the opportunity. Preference reversal may, then, be taken into account in the sense that had the non-receiving agent accepted the opportunity she would also have obtained the pay raise, and as such regret as a counterfactual belief based emotion might change a undesired action into a desired one, however not because the non-receiving, necessarily, now desires to accept the possibility offered, rather she may solely regret declining the opportunity which could have brought her the desired reward. In other words, the non-receiving may not want to change her action, just the result. In a similar manner, if no other cognitive scenario can be produced leading to a better result in which she cannot blame herself, she must accept that she only has herself to blame, see also chapter III.

Comparing case-1 with case-3 seems to show that the difference in odds is fairly small between the respondents finding that no norm violation had occurred in a case with pure exogenously produced differences to one representing a mix is rather small. In case-3 the odds are 53.8 compared to 38.8 in case-1. This might be compared with the odds of finding a norm violation to have occurred, although the results are not statistically significant, in case-3 the odds are 0.1 and in case-1 the odds are 0.5. It might, in other words, be speculated that purely exogenously produced differences lowers the possibility of assigning horizontal blame within horizontal relations. In accordance with what was argued in chapter VI this might be due to the lack of the possibility of assigning causal agency. Against this, it might be argued that the majority respondents in neither of the cases found a norm violation to have occurred. On case-1, one of the respondents argues:

---

<sup>44</sup> Filename: Organization X, time code: 00:06:17

*Interviewee:* The first guy [*Peter*] did not do anything. It was not his fault. It was someone else who decided that the other guy [*Søren*] should not participate in the...what did you call it?

*Interviewer:* Project group.

*Interviewee:* Yes.

*Interviewer:* Would it have made a difference?

*Interviewee:* I do not know... Well, he [*Søren*] would not have gotten anything. But the other...yeah, he might have felt that it was his fault.”<sup>45</sup>

On case-3, the respondent argues:

*Interviewer:* Earlier we spoke about that it might perhaps be easier to accept if it came from the outside, as it does in this story. Is it easier for him [*Jørgen*] to accept that he was not offered the opportunity, but the other one was?

*Interviewee:* I would mean so. But the other one [*Jørgen*] does not sound like he would be interested.

*Interviewer:* What about the difference itself?

*Interviewee:* Well, perhaps he should have been offered something else.”<sup>46</sup>

Based on this it might be argued that exogenously produced compensation differences are easier to accept, just as they might be less threatening to the receiving agent's self-perception – that is, it might be easier for this agent to convince herself that it was not her actions which made the non-receiving agent worse off. What is interesting, however, might be the fact that indirect causal agency does not seem to liberate the social agent from feeling guilty of having produced these externalities. This seems to imply that even though the agent does not produce these externalities directly through his actions, insofar as he cannot decide whether he is rewarded or not does not limit his possibility of feeling guilty. Additionally, it should be noticed that indirect agency may not be perceived as having a vertical influence, in the sense that it will have no influence on the results. Thus, the action tendencies described by equity theorists (chapter V) may solely be applied to restore the agent's private sense of fairness or to restore horizontal relations, but it will not have any vertical influence due to the causal nature of the re-

---

<sup>45</sup> Filename: Organization X, time code: 00:01:58

<sup>46</sup> Filename: Organization X, time code: 00:10:38

ward process. Moreover, preferences may also be taken into consideration, insofar as a reward which might represent no significance to the other social agent may not produce externalities associated with the reward itself, but only the difference that it comes to represent. It might, of course, in relation with what was argued in chapter VIII be discussed whether the emotion invoked in the situation is actually guilt or shame. In accordance with Elster's (1999) distinction, discussed in chapter VIII, it might be argued that insofar as a social norm existed on causal agency, the agent would feel shame and not guilt. The respondent in the interview, however, relates causal agency with a kind of unconditional moral norm stating that one is not allowed upset one another – one should always act appropriately. Implied in this notion of appropriate behavior seems to a kind of social rationality related with how agents should interact with one another. The normative should which seems to be implied in the interviewee's argument may take on moral normative nature meaning that in situations when no norm transgression has occurred, the agent is still morally responsible for acting in an appropriate manner. In addition to this, the interviewee argues with regard to case-1:

*Interviewee:* ...he [Peter] should of course not brag about it.

*Interviewer:* Why not?

*Interviewee:* You just do not do that.

*Interviewer:* Is that an unwritten rule?

*Interviewee:* No, I would call it appropriate behavior.

*Interviewer:* He, who receives the bonus, is he doing something which is against this appropriate behavior?

*Interviewee:* What do you mean?

*Interviewer:* By choosing to go for the bonus?

*Interviewee:* No, he is allowed to do that...but of course if he knew that it meant a lot to the other guy.

*Interviewer:* He should not have done it?

*Interviewee:* No, but it might ruin something."<sup>47</sup>

---

<sup>47</sup> Filename: Organization X, time code: 00:02:31

Later in continuation of this, the interviewee argues:

*“Interviewee:* ...you do not want to share an office with someone who does not like you. What I mean is just that it might ruin something...it might ruin the good mood. No one likes to be around someone who does not like them.

*Interviewer:* So there are limits to how ambitious one is allowed to be?

*Interviewee:* No, but you just have to make it work.”<sup>48</sup>

Based on this, it might be argued that there are no actual limits preventing the receiving agent from pursuing some desired end, at least not as long as he does it within the boundaries of appropriate behavior. These boundaries of appropriateness concerns actions which might make other social agents worse off. These rules, however, should not be perceived as limiting professional behavior; rather they should be perceived as purely social, insofar as they are related with the social life of horizontal relations. Of course, this might only apply in professional relations, in the sense that close and personal relations might be governed by other rules.

These rules, as such, are not related with the distribution of goods, but primarily have as their target the behavior of the social agents, although they indirectly may, at least to some extent, also come to govern choices regarding which goals pursue, if the agent, for example, knows that her actions might make another agent worse off. In other words, this seems to imply that the social agent must balance her goals and desires with those held by other social agents and, although, it should not prevent her from being ambitious, some goals might be considered as violating these rules. Social norms governing behavior may, then, exist as boundaries on which actions to choose, in the sense that they aim at promoting pro-social behavior. According to this, the agent should comply with these social norms due to relational concerns, in the sense that she might create unnecessary conflicts, just as she might adhere to these because she prefers being liked. The interviewee, it seems, provide three different reasons: (1) it might make the other social agent worse off; (2) relational concerns or belongingness; and (3) esteem/self-esteem. These reasons seem, at least to some extent, to be equivalent with the ones proposed in the previous chapters. Moreover, the quotation also seems to illustrate the notion of social rationality as a principle of social action aiming primarily at securing the continuation of social relations and to produce actions in accordance with gov-

---

<sup>48</sup> File name: Organization X, time code: 00:03:40

erning social and moral norms. Social rationality is not, as such, in opposition to ambition, rather it guides the agent in how she should pursue her will in a socially rational manner, see also chapters III and VIII.

Figure 42

**Emotional Value of Receiving: "Please evaluate how much you believe the reward to mean to the receiving agent?"**

	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative percent</i>	<i>Odds ratio with middle category as base-line</i>	<i>90% Confidence intervals (Lower bound, Upper bound)</i>
<b>Case-1</b>	Negative	3	1	1	0.05 (0.02, 0.13)
	Neutral	36	17	18	1.0*** (0.68, 1.48)
	Positive	176	82	100	22.2*** (16.45, 30.08)
	Total	215	100		
<b>Case-2</b>	Negative	0	0	0	-
	Neutral	19	10	10	1.0*** (0.59, 1.71)
	Positive	178	90	100	81.0*** (54.39, 120.62)
	Total	197	100		
<b>Case-3</b>	Negative	0	0	0	-
	Neutral	6	3	3	1.0 (0.39, 2.59)
	Positive	184	97	100	1045.4*** (527.26, 2072.9)
	Total	190	100		
	Peter and Søren	128			
	Pemille and Jeanette	146			
Missing data	Anders and Jørgen	153			

From this discussion of horizontal norms and vertical fairness, I now turn to the emotional significance of receiving. The items on the 9 point Likert-scale have been collapsed into three categories: the first category covers “high negative significance” 1 to 3; the second category covers the items 4 to 6; and the last category covers the items 7 to 9 “highly positive significance”. Across the cases, it seems that the majority of the respondents tend to find the emotional value to be positive. In case-1, the odds of finding the bonus to represent a positive emotional value to judging it to be neutral are 22.2. In case-2, on the other hand, the odds finding it to represent a positive value are 81. In case-3, the odds of finding it to represent a positive value are 1045.4. The differences in the odds of finding it to represent a positive value, across the cases, might be due to either expectations or the nature of the reward. In the first two cases, the receiving agent does something to obtain the reward; hence he/she expected it, while in the third case the agent did not do something in particular to bring the after-situation about – no direct agency can be identified, at least not one which is purposively employed by the agent in order to bring about a particular after-situation. Although, the receiving agent in the third vignette has done something to bring about the situation, the reward may be perceived as future directed and, as such, it might come to, for example, represent apprecia-

tion, whereas getting rewarded for actions already performed might, in somewhat different sense, signal *just* being paid. Hence, backward directed rewards may not represent the same emotional significance as future directed ones; as such this implies a distinction between *ex ante* and *ex post* reward policies (Lawler, 2000). Another explanation might be that educational opportunities are more highly valued than monetary compensation. The latter might be explained as the symbolic meaning expressed in an educational opportunity, insofar as it signifies that the organization is aiming at establishing a long term relation with the employee (Schuster and Zingheim, 2000; see also chapter II), just as it might, simply, reflect the preferences of the respondent group.

Figure 43

**Emotional Value of Not Receiving: "Please evaluate how much you believe the compensatory difference to mean to the non-receiving individual?"**

	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative percent</i>	<i>Odds ratio with middle category as base-line</i>	<i>90% Confidence intervals (Lower bound, Upper bound)</i>
<b>Case-1</b>	Negative	95	45	1.7***	(1.34, 2.20)
	Neutral	68	32	1.0***	(0.75, 1.33)
	Positive	49	23	0.6***	(0.47, 0.86)
	Total	212	100		
<b>Case-2</b>	Negative	73	37	0.6***	(0.47, 0.79)
	Neutral	97	49	1.0***	(0.79, 1.27)
	Positive	28	14	0.2	(0.12, 0.24)
	Total	198	100		
<b>Case-3</b>	Negative	94	49	2.0***	(1.56, 2.68)
	Neutral	61	32	1.0	(0.74, 1.35)
	Positive	36	19	0.5***	(0.35, 0.71)
	Total	191	100		
	Peter and Søren	131			
	Pemille and Jeanette	145			
Missing data	Anders and Jørgen	152			

From the emotional value of receiving, I turn to the emotional value of not receiving. As the table demonstrates, the respondents' answers seem more distributed between the categories. In case-1, the odds of finding the compensatory difference to represent a negative emotional value to the non-receiving individual are 1.7. In case-3, the odds of finding it to represent a negative value are 2, while the odds of finding it to represent a negative value in case-2 are 0.6. Intuitively, it seems that the likelihood of finding it to represent a negative value to the non-receiver in case-1 and in case-3 is slightly higher than finding it to be neutral, just as the likelihood of finding it to represent a positive values is less likely than finding it neutral. Whereas in case-2, the odds of finding it to represent a neutral value are slightly higher, this might be interpreted as a result of the

fact that the non-receiver declined the offer. Another interesting result here, is also that the odds of finding it to represent a negative value in case-1 is not as high as one might have predicted, because the non-receiving individual tried to become a member of project team himself. Although, the distribution of the answers varies across the three cases, the variation is too small to be statistically significant; hence, the comparisons made here represent nothing but speculative interpretations.

From a social comparison perspective, it might be argued that in cases 1 and 2, the non-receiving agent could not expect being rewarded, insofar as the compensatory difference follows from an action made by the receiving agent. Relative deprivation would also predict that the negative emotional experience in case-3 would be more negative due to fact that the relational structure shifted from one characterized by equality to characterized by inequality. On case-1, one of the respondents argues:

“Interviewer: What do you think the pay difference means to Søren?”

Interviewee: What do you mean?

Interviewer: Positive or negative?

Interviewee: Negative, I would think.

Interviewer: Why?

Interviewee: It is never fun to be in that situation.

Interviewer: What do you mean?

Interviewee: Well, to be the one who does not get anything.

Interviewer: Do you think it have any significant influence that Søren used to get more than Peter?

Interviewee: Perhaps, but then it would not be the entire reason (...) if such a matter bears any significance, it is about something between them. But of course if he is used to getting the most, then...well...it might feel a bit strange. One might feel that one does not have the same value as before.”<sup>49</sup>

In the same line of reasoning, the interviewee argues on case-2:

“Interviewer: What about her [*Jeanette*] who did not get anything?”

Interviewee: Again, I think she would feel less appreciated.

---

<sup>49</sup> Filename: Organization Y, time code: 00:03:16

Interviewer: So it makes no difference which one who used to get the most?

Interviewee: Perhaps, but then it is because one no longer feels as appreciated as before.<sup>50</sup>

Based on this it might be argued that the horizontal difference, in itself, bears a symbolic significance as a sign on vertical appreciation. Relational differences before and after, then, can primarily be viewed as changes in vertical appreciation, the difference might matter, but then it shifts to the horizontal relation between the receiver and the non-receiver. The vertical perspective developed here, seems to be in accordance with Van den Bos, Lind, Vermunt, and Wilke's (1997) perspective and the relational approach to organizational fairness, while the horizontal perspective seems to be in accordance with the perspective developed by Fehr and Schmidt (1999), see chapters III and V. Moreover, it also seems to explicate the difference between the emotional foundation of the vertical and horizontal relations, insofar as the vertical relation is centered upon emotions associated with appreciation, while the horizontal perspective is associated with social relational emotions, although the Fehr-Falk model also introduces this in vertical relations as previously argued in this chapter, see chapters III and IV. From what is argued here, the change from the before- to the after-situation may matter because it takes on a different symbolic meaning. However, it should be noted that the loss of appreciation does not follow from a decrease in compensation, but rather from a loss of horizontal position, indicating that these positions bear vertical symbolic significance, whereas from a horizontal perspective it might not matter who receives most unless something inherent to their relation makes it significant. The emphasis on non-material rewards can be related with the socioemotional structuring of the relation and the Humean perspective and, as such, from this perspective it might be more important to be liked (included) and treated fairly than being the highest paid. The desire to belong may, of course, also manifest itself in vertical relations as argued by, among others, Haidt (2012) in chapter VIII. In accordance with what was argued above, the less symbolic significance of status differences in horizontal relations may be associated with the necessity of behaving in a social rational manner.

Furthermore, the lack of variation across the cases might be interpreted as an argument against the impact of relational status quo and hedonic status quo, insofar as one

---

<sup>50</sup> Filename: Organization Y, time code: 00:09:22

would expect the negative emotional value in case-2 and case-3 to be higher, because in the first case, the non-receiving individual goes from being above to being below, and in the latter case, the non-receiving individual goes from being equal to being below in terms of compensatory outcome. Combining the hedonic perspective with the relational status quo perspective, it should be expected that these changes would be framed as a loss to the non-receiving individual and, as such, be associated with a higher negative emotional influence on the non-receiver. Again, the variation across the cases is not statistical significant, hence the tendency cannot be confirmed.

Figure 44

**Relational Influence: "Please evaluate how the new situation will influence the relation between the receiving and non-receiving agent?"**

	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative percent</i>	<i>Odds ratio with middle category as base-line</i>	<i>90% Confidence intervals (Lower bound, Upper bound)</i>	
<b>Case-1</b>	Negative	63	30	30	0.4***	(0.34, 0.58)
	Neutral	105	49	79	1.0***	(0.80, 1.26)
	Positive	45	21	100	0.3	(0.21, 0.37)
	<b>Total</b>	<b>213</b>	<b>100</b>			
<b>Case-2</b>	Negative	65	33	33	0.4***	(0.32, 0.54)
	Neutral	105	54	87	1.0***	(0.80, 1.26)
	Positive	26	13	100	0.1	(0.09, 0.18)
	<b>Total</b>	<b>196</b>	<b>100</b>			
<b>Case-3</b>	Negative	73	38	38	0.8***	(0.65, 1.10)
	Neutral	81	42	81	1.0	(0.77, 1.30)
	Positive	37	19	100	0.3	(0.23, 0.45)
	<b>Total</b>	<b>191</b>	<b>100</b>			
	Peter and Søren	130				
	Permille and Jeanette	147				
Missing data	Anders and Jørgen	152				

From this discussion of emotions, I now turn to the relational influence of the compensatory difference. The 9 point Likert-scale have here been collapsed into three categories, the first category ranges from “High negative influence” 1 to 3, the second category from 4 to 6, and the last category from 7 to 9 “High positive influence”. Across all the cases, it seems that the majority respondents find the relational influence or change to be either negative or neutral. In case-1, the odds of finding it to represent a negative change are 0.4 to finding it neutral. In case-2, the odds of finding it to be negative are 0.4. In case-3, the odds of finding it to represent a negative change are 0.8. The tendency across all the cases, then, seems to be that the respondents are more likely to find it to represent a neutral value.

In section 10.5.4, I will further discuss the relation between the different variables and relational influence. From what have been argued so far in the interviews, it seems that the relational influence might be a matter of the nature of the horizontal relation, just as it might be related with other matters such as the emotional significance of the reward and the compensatory difference to the receiving and non-receiving agents.

Figure 45

Summary table of the relation between the independent and dependent variables

Case variables/independent variables		Dependent variables		Odds ratio with middle category as base-line	90% Confidence intervals (Lower bound, Upper bound)		
Group of variables	State	Variable name	State				
<b>Peter and Søren case</b>	<i>Relational variables</i>	Educational background	Equal	Distributive fairness	Unfair	0.3	(0.17, 0.66)
			Equal	Neutral	1.0***	(0.61, 1.63)	
			Equal	Fair	49.7***	(34.51, 71.58)	
	Job title	Equal	Unfair	0.3	(0.14, 0.46)		
		Equal	Neutral	1.0***	(0.68, 1.48)		
		Equal	Fair	17.3***	(12.78, 23.44)		
	Competence level	Equal	Procedural fairness	Unfair	0.3	(0.14, 0.46)	
				Neutral	1.0***	(0.68, 1.48)	
				Fair	17.3***	(12.78, 23.44)	
	Seniority	Unequal, non-receiving > receiving	Norm violation	No norm violation	38.5***	(27.23, 54.43)	
				Neutral	1.0***	(0.63, 1.58)	
				Norm violation	0.3	(0.16, 0.58)	
	<i>Compensation</i>	Before	Unequal, non-receiving > receiving	Negative	0.05	(0.02, 0.13)	
				Neutral	1.0***	(0.68, 1.48)	
				Positive	22.2***	(16.45, 30.08)	
	After	Equal	Emotional value of receiving	Negative	1.7***	(1.34, 2.20)	
				Neutral	1.0***	(0.75, 1.33)	
				Positive	0.6***	(0.47, 0.86)	
Intentionality	Yes	Emotional value of not receiving	Negative	0.4***	(0.34, 0.58)		
			Neutral	1.0***	(0.80, 1.26)		
			Positive	0.3	(0.21, 0.37)		
<b>Pernille and Jeanette case</b>	<i>Relational variables</i>	Educational background	Equal	Distributive fairness	Unfair	0.2	(0.08, 0.37)
			Equal	Neutral	1.0***	(0.65, 1.54)	
			Equal	Fair	25.8***	(18.51, 16.00)	
	Job title	Equal	Unfair	0.1	(0.07, 0.29)		
		Equal	Neutral	1.0***	(0.67, 1.48)		
		Equal	Fair	17.1***	(12.59, 23.33)		
	Competence level	Equal	Procedural fairness	Unfair	0.1	(0.07, 0.29)	
				Neutral	1.0***	(0.67, 1.48)	
				Fair	17.1***	(12.59, 23.33)	
	Seniority	Unequal, non-receiving > receiving	Norm violation	No norm violation	776.0***	(367.46, 1638.77)	
				Neutral	1.0	(0.35, 2.84)	
				Norm violation	0.7	(0.20, 2.20)	
	<i>Compensation</i>	Before	Unequal, non-receiving > receiving	Negative	-	-	
				Neutral	1.0***	(0.59, 1.71)	
				Positive	81.0***	(54.39, 120.62)	
	After	Unequal, non-receiving < receiving	Emotional value of receiving	Negative	0.6***	(0.47, 0.79)	
				Neutral	1.0***	(0.79, 1.27)	
				Positive	0.2	(0.12, 0.24)	
Intentionality	Yes	Emotional value of not receiving	Negative	0.4***	(0.32, 0.54)		
			Neutral	1.0***	(0.80, 1.26)		
			Positive	0.1	(0.09, 0.18)		
<b>Anders and Jørgen case</b>	<i>Relational variables</i>	Educational background	Equal	Distributive fairness	Unfair	0.2	(0.07, 0.46)
			Equal	Neutral	1.0	(0.59, 1.71)	
			Equal	Fair	66.0***	(44.26, 98.41)	
	Job title	Equal	Unfair	0.2	(0.07, 0.36)		
		Equal	Neutral	1.0	(0.65, 1.53)		
		Equal	Fair	23.9***	(17.22, 33.23)		
	Competence level	Equal	Procedural fairness	Unfair	0.2	(0.07, 0.36)	
				Neutral	1.0	(0.65, 1.53)	
				Fair	23.9***	(17.22, 33.23)	
	Seniority	Equal	Norm violation	No norm violation	53.8***	(36.99, 78.18)	
				Neutral	1.0	(0.61, 1.64)	
				Norm violation	0.1	(0.01, 0.40)	
	<i>Compensation</i>	Before	Equal	Negative	-	-	
				Neutral	1.0	(0.39, 2.59)	
				Positive	1045.4***	(527.26, 2072.9)	
	After	Unequal, non-receiving < receiving	Emotional value of receiving	Negative	2.0***	(1.56, 2.68)	
				Neutral	1.0	(0.74, 1.35)	
				Positive	0.5***	(0.35, 0.71)	
Intentionality	No	Emotional value of not receiving	Negative	0.8***	(0.65, 1.10)		
			Neutral	1.0	(0.77, 1.30)		
			Positive	0.3	(0.23, 0.45)		

### 10.5.2 Fairness and Norm Violation

The aim of this subsection is to study the relation between fairness, distributive and procedural, and norm violation. From what has been discussed in the previous chapters, finding a difference vertically fair may not necessarily be the same as finding it horizontally fair – that is, as not being indirectly produced by a norm-violating action endogenous to the horizontal relation. Initially, consider the correlation matrix below:

Figure 46

**Correlation matrix**

	<i>Distributive fairness</i>	<i>Procedural fairness</i>	<i>Norm violation</i>	<i>Emotions of receiving</i>	<i>Emotions of non-receiving</i>	<i>Relational influence</i>
<b>Case-1</b>	Distributive fairness	1,000	,657**	-,423**	0,049	0,062
	Procedural fairness	,657**	1	-,485**	,144*	0,057345948
	Norm violation	-,423**	-,485**	1	-0,084289302	-0,064291844
	Emotions of receiving	,049	,144*	-,084	1,000	-,148*
	Emotions of non-receiving	,062	,057	-,064	-,148*	1,000
	Relational Influence	,105	,182**	-,081	-,016	,321**
<b>Case-2</b>	Distributive fairness	1,000	,622**	-,434**	,232**	,236**
	Procedural fairness	,622**	1	-,456**	,214**	,271**
	Norm violation	-,434**	-,456**	1	-,342**	-0,045423511
	Emotions of receiving	,232**	,214**	-,342**	1	0,114224982
	Emotions of non-receiving	,236**	,271**	-0,04542	0,114224982	1
	Relational Influence	,241**	,248**	-0,08599	-0,025871979	,478**
<b>Case-3</b>	Distributive fairness	1,000	,717**	-,752**	,478**	0,097
	Procedural fairness	,717**	1	-,679**	,406**	,167*
	Norm violation	-,752**	-,679**	1	-,332**	-,187**
	Emotions of receiving	,478**	,406**	-,332**	1	-0,014876351
	Emotions of non-receiving	,097	,167*	-,187**	-,015	1,000
	Relational Influence	,214**	,271**	-,266**	-0,009879317	,591**

\* P ≤ 0.05  
\*\* P ≤ 0.01

Across all the cases, distributive fairness is fairly positively correlated with procedural fairness. This observation is in accordance with the overall fairness perspective on organizational fairness, see chapter III. Moreover, across all the cases, the norm violation variable is inversely correlated with the two measures of fairness. In case-1 and in case-2, the correlation between the variable “Norm violation” and the two measures on fairness is rather low, whereas in case-3 it is fairly inversely correlated. That is, when the two fairness measures go toward fairness, the norm violation goes towards no norm violation despite the fact that fairness, primarily, is a vertical phenomenon and norm violation, as defined here, a horizontal phenomenon. Outcome claims may, of course, as argued in chapters V and VII be horizontally defensible, insofar as they are related with

the relational and temporal structuring of the claim, while the norm violation is related with the behavioral dimension.

To get a more thorough understanding of the relation between fairness and the “Norm violation” variable, I will now turn to an analysis of the relation in cases 1 and 3.

Figure 47

**Frequency and Percentage Crosstabulations of the Relation between Distributive Fairness and Norms on Taking**

		Violation of a norm on receiving			
		No norm violation	Neutral	Norm violation	Total
<b>Case-1</b>	Frequency Crosstabulation				
	Unfair	20	10	1	31
	Fair	159	16	8	183
	Total	179	26	9	214
<b>Case-1</b>	Percentage Crosstabulation				
	Unfair	11%	38%	11%	14%
	Fair	89%	62%	89%	86%
	Total	100%	100%	100%	100%
<b>Case-3</b>	Frequency Crosstabulation				
	Unfair	12	10	1	23
	Fair	154	12	0	166
	Total	166	22	1	189
<b>Case-3</b>	Percentage Crosstabulation				
	Unfair	7%	45%	100%	12%
	Fair	93%	55%	0%	88%
	Total	100%	100%	100%	100%

	Significance level	Missing data
Peter and Søren case	0.002	129
Anders and Jørgen case	0.074	146

First of all, it should be noted that the “Distributive fairness” variable have been collapsed into two categories, the first category covers the items 1 to 6, while the second category covers the items 7 to 9. This has been done in order to better capture the effect of differences. Furthermore, it have been chosen only to study the relationship between the variable “Norm violation” and the variable “Distributive fairness”, because the two measures are fairly positively correlated and the variable “Norm violation” is fairly sim-

ilarly correlated to both fairness measures, hence studying the relation between both measures of fairness and “Norm violation” seems redundant.

In case-1, the odds ratio of finding the reward to be distributive fair when simultaneously finding that no norm violation has occurred relative to finding that a norm violation has occurred is 1.01 indicating close to no relationship, while the odds ratio of finding the reward fair when simultaneously finding that no norm violation has occurred relative to holding a neutral position on whether a norm violation has occurred is 0.2 [ $p < 0.1$ ; 90% CI (0.093, 0.435)] indicating that the likelihood of finding the reward to be distributive fair is higher if one also finds that no norm violation had occurred compared to holding a neutral opinion on whether a norm violation had occurred. That is, one is more likely to find the reward fair than unfair if one simultaneously find that no norm violation had occurred compared with holding a neutral opinion. In case-3, a similar tendency seems to found, although the odds ratio cannot be calculated because 0 found that a norm violation had occurred and answered that it was highly distributive fair, on the other hand, the odds ratio of finding the reward fair to finding it unfair while also finding that no norm violation had occurred compared to the neutral opinion is 0.09 [ $p < 0.1$ ; 90% CI (0.039, 0.221)] indicating that the likelihood of finding it fair while also finding that no norm violation has occurred is higher than if on holds a neutral opinion.

*Result-1:* Norm violation is inversely correlated with fairness implying that when the variable “Norm violation” goes towards no norm violation, fairness goes towards more fair. Furthermore, this tendency also entails that the likelihood of finding the outcome distribution to be fair increases if it is judged that no horizontal norm violation have occurred or at least compared with holding a neutral opinion on the matter. It should be noted that because very few respondents found the outcome distribution unfair and that a norm violation had occurred, hence it might be problematic to state that this relation also holds in cases with vertical unfairness and horizontal norm violations.

From what was discussed in the previous section, fairness is primarily a vertical phenomenon, while norm violations are horizontal in nature. Fairness may be associated with the actions of the receiving agent, in the sense that she/he has done something to bring the after-situation about, and by bringing this new state about, the receiving agent might have violated some horizontal social norm – although it might be argued that due to lack of causality, the norm is targeting the action and not the actual after-situation in

itself. However, quasi moral norms on appropriate behavior which seems to govern horizontal relations may also influence the fairness judgment, insofar as the after-situation, as already discussed, may be indirectly brought about by the actions of the receiving agent, or the non-receiving agent as in case-2, and directly brought about by some external decision. Hence, what is argued, here, is that fairness judgments may be influenced by horizontal judgments related with whether or not the agents violated some norm and vertical fairness judgments related with the causal process which brought the situation about. It is, of course, problematic to postulate a causal relation between the two variables, so it might simply be argued that the chance of finding the outcome fair from a distributive perspective increases if the respondent also finds that no norm violation has occurred. This might be interpreted as the likelihood of finding a compensatory difference to be fair increases if it is also judged that this was indirectly and directly brought about by fair actions, and as representing principles which are in accordance with the agent's own or that the agent, at least, accept these – meaning that she is not able to provide any strong reasons against them, see chapter III.

### **10.5.3 Fairness, Norm Violation, and Emotions**

In the previous section, I focused on the relation between fairness judgments and horizontal norm violation. In this section, I turn to the relation between fairness, norm violation, and the emotional significance of the compensatory difference to the receiving and non-receiving individual. Although, the distribution of answers does not allow an analysis of variation across the cases, it is possible to study the general tendencies characterizing these relations.

In figure 46, the correlation between emotions, norms, and the two fairness measures is showed. Emotions are only weakly correlated with these three other variables. However, based on the theories discussed in chapter III, it might be argued that fairness should be positively associated with the emotions invoked in the receiving individual, insofar as the reasonable individual would gain most pleasure from a fair distribution.

Figure 48

**Frequency and Percentage Crosstabulations of the Relation between Distributive Fairness and the Emotional Value of Receiving**

Distributive fairness	Emotional value of receiving			
	Negative	Neutral	Positive	Total
<b>Case-1</b>				
Frequency Crosstabulation				
Unfair	2	3	26	31
Fair	1	33	150	184
Total	3	36	176	215
Percentage Crosstabulation				
Unfair	67%	8%	15%	14%
Fair	33%	92%	85%	86%
Total	100%	100%	100%	100%
<b>Case-3</b>				
Frequency Crosstabulation				
Unfair	0	2	20	22
Fair	0	4	162	166
Total	0	6	182	188
Percentage Crosstabulation				
Unfair	0%	33%	11%	12%
Fair	0%	67%	89%	88%
Total	100%	100%	100%	100%

	Significance level	Missing data
Peter and Søren case	0.041	128
Anders and Jørgen case	0.094	155

In case-1, the odds ratio of finding the distributive fair to unfair, while also finding the emotional value to the receiving individual to be neutral relative to finding it to represent a positive value is 0.52 [ $p > 0.1$ ; 90% CI (0.183, 1.506)]. In case-3, the odds ratio is 4.05 [ $p > 0.1$ ; 90% CI (0.921, 17.818)]. None of the relations are statistically significant, but points towards the trend that the likelihood of finding the outcome distribution unfair is higher if one also finds the emotional value of the reward to be neutral compared to finding the value positive.

From this, I turn to the relation between distributive fairness and the emotions of not receiving. Consider the table below:

Figure 49

**Frequency and Percentage Crosstabulations of the Relation between Distributive Fairness and the Emotional Value of not Receiving**

Distributive fairness	Emotional value of not receiving			
	Negative	Neutral	Positive	Total
<b>Case-2</b>				
Frequency Crosstabulation				
Unfair	20	10	5	35
Fair	53	87	22	162
Total	73	97	27	197
Percentage Crosstabulation				
Unfair	27%	10%	19%	18%
Fair	73%	90%	81%	82%
Total	100%	100%	100%	100%
Significance level	0.015			
Missing data	146			

From the figure it seems that the majority of the respondents find outcome difference to represent an emotional neutral value to the non-receiving individual. The odds ratio of finding the compensatory difference to be fair while also judging it to represent a neutral value relative to finding it to represent a negative value is 3.28 [ $p < 0.1$ ; 90% CI (1.629, 6.616)], while the odds ratio of finding it fair when also judging it to represent a neutral value relative to finding it to represent a positive value is 0.51 [ $p > 0.1$ ; 90% CI (0.189, 1.355)], and the odds ratio of finding it unfair to fair while also the emotional value negative relative to finding it positive is 1.66 [ $p > 0.1$ ; 90% CI (0.658, 4.188)]. From this it seems to follow that one is more likely to find the compensatory difference unfair to fair if one also finds the value negative relative to finding it neutral – or equivalent one is more likely to find it fair to unfair if one also finds the emotional value neutral compared to if one finds the value negative. This case is, of course, special because the non-receiving individual declined the opportunity and, as such, this might explain why the respondents find it to represent a neutral value. Furthermore, it is also interesting that even though the respondents find the compensatory difference to represent a neutral value, they also judge it to be fair. From a first-person perspective one might expect that envy may bias the judgment – making a fair difference seem unfair, just as the hedonic experience accompanying a relational change from above to below might drive the judgment towards unfairness, see chapters III, V, and VII. However, from a

third-person perspective, the emotional experience might be decoupled from the fairness judgment – therefore, enabling a more rational judgment in which negative emotions associated with not receiving coexist with fairness. Causal agency on behalf of the non-receiving individual might also explain it, insofar as the agent who declines a possibility cannot blame anyone but herself.

*Result-2a:* From the perspective of the receiving individual, it seems that the likelihood of finding the distribution of the outcome to be fair if she also finds it to represent a positive emotional value is greater than if she found it neutral, although the result was not statistical significant. From the perspective of the non-receiving individual, it seems that the likelihood of finding a compensatory difference fair is greater if one also finds it to represent a neutral emotional value. Although, this latter statement should be viewed in relation with causal agency on behalf of the non-receiving individual – the non-receiving individual had a choice – and expressed preferences associated with causal agency.

Figure 50

**Frequency and Percentage Crosstabulations of the Relation between the Emotional Value of not Receiving and Norms on Taking**

Emotional value, not receiving	Violation of a norm on receiving			
	No norm violation		Norm	
	violation	Neutral	violation	Total
<b>Case-1</b>				
Frequency Crosstabulation				
Negative	78	13	4	95
Neutral	65	3	0	68
Positive	35	9	5	49
Total	178	25	9	212
Percentage Crosstabulation				
Negative	44%	52%	44%	45%
Neutral	37%	12%	0%	32%
Positive	20%	36%	56%	23%
Total	100%	100%	100%	100%
<b>Case-2</b>				
Frequency Crosstabulation				
Negative	72	1	0	73
Neutral	95	1	1	97
Positive	23	3	2	28
Total	190	5	3	198
Percentage Crosstabulation				
Negative	38%	20%	0%	37%
Neutral	50%	20%	33%	49%
Positive	12%	60%	67%	14%
Total	100%	100%	100%	100%

	Significance level	Missing data
Peter and Søren case	0.04	131
Pernille and Jeanette	0.04	145

From the relation between the emotional value of not receiving and fairness, I now turn to the relation between the emotional value of not receiving and the “Norm violation” variable. In case-1, the odds ratio of finding the emotional value of not receiving to be negative to neutral, while also finding that no norm violation has occurred, relative to holding a neutral position on whether a norm violation has occurred is 0.28 [ $p < 0.1$ ; 90% CI (0.093, 0.826)], whereas the odds ratio of finding it to represent a positive emotional value to a neutral, while also finding that no norm violation has occurred, relative to the neutral position is 5.57 [ $p < 0.1$ ; 90% CI (1.759, 17.651)]. These results indicate that one is more likely to find the emotional value negative if one also holds a neutral position on whether a norm violation has occurred compared to finding that no norm violation had occurred, just as one is more likely to find the emotional value neutral to positive if one also finds that no norm violation had occurred relative to a neutral opinion. Hence, even when one finds that the receiving agent has not violated any social norms, one is more likely to find the compensatory difference to represent a neutral emotional value than a positive one. Thus, in vertical relations the compensatory difference is more likely to be found fair if it represents a neutral emotional value to the non-receiving value, while if this compensatory difference, from a horizontal perspective, is not the outcome of norm violating actions, the non-receiving agent is more likely to find it to represent a neutral emotional value.

This might be interpreted as stating that compensatory differences brought about by non-norm violating actions might still produce negative emotions within the non-receiving individual, although these might not be directly related with the actions of the receiving agent – a fact which is also visible in the low correlation between the two variables. That is to say, these emotions may not be related with a possible norm transgression, but primarily with other matters. In the present case, the non-receiving individual tried to become a member of the project team, but was rejected; hence he might not be able to blame the receiving individual directly, although indirectly he may argue to himself that if the receiving agent had not become a member, he could have been. As such, this may be related with the non-receiving agent’s possibility of producing counterfactual conclusions, see chapter VI.

In case-2, the odds ratio of finding the emotional value of the compensatory difference to represent a negative value to a neutral one, while finding that no norm violation had occurred, relative to a neutral position is 0.76 [ $p > 0.1$ ; 90% CI (0.072, 7.928)], the odds ratio of finding it to be positive to neutral, while finding that no norm violation has occurred relative to holding the neutral opinion is 12.39 [ $p < 0.1$ ; 90% CI (1.775, 86.527)], and the odds ratio of finding it to represent a negative value to a positive one, while finding that no norm violation had occurred relative to holding a neutral position is 9.39 [ $p < 0.1$ ; 90% CI (1.342, 65.733)]. These results seem to indicate that one is more likely to finding the emotional value of the compensatory difference to be neutral than positive if one also finds that no norm violation had occurred relative to holding the neutral opinion, just as one is more likely to find it to represent a negative emotional value to a positive one if one finds that no norm violation had occurred relative to the neutral position.

Again, it can be argued that the negative emotional value produced, here, is not a result of norm transgression on behalf of the receiving agent; rather the emotions might be related with other matters. Although, here it might be argued that the non-receiving agent declined the opportunity offered to her which might imply that the negative emotions are related with the difference itself, because these may not be related with the underlying reward, insofar as the non-receiving declined her opportunity, just as these cannot be, according to the respondents' judgment, related with a norm transgression. It might be speculated that the emotional value is related to the relational status quo, in the sense that the non-receiving individual goes from being above to being below. However, this cannot be statistically confirmed due to the lack of variation in response distributions across the cases.

*Result- 2b:* In cases, when no norm-violation has occurred, the negative emotions associated with a compensatory difference may either be due to cognitive normative expectations and the possibility of producing counterfactual conclusions, or in cases when no norm violation has occurred and the non-receiving agent has revealed her preferences and through causal agency declined an opportunity, the negative emotions might, although it can only be speculated, be related with relational status quo and hedonic adaption – or simply, as argued above, representing that the non-receiving can only blame herself.

### 10.5.4 Relational Influence

In this section, I turn to an exploration of the relational influence dimension. Here, I will consider relational influence as the dependent variable and explore its relation with the fairness measures, the norm violation variable, and the emotion variables.

Figure 51

**Frequency and Percentage Crosstabulations of the Relation between Relational Influence and Distributive Fairness**

Relational Influence	Distributive Fairness			
	Unfair	Neutral	Fair	Total
<b>Case-1</b>				
Frequency Crosstabulation				
Negative	2	6	22	30
Neutral	3	16	136	155
Positive	3	1	24	28
Total	8	23	182	213
Percentage Crosstabulation				
Negative	25%	26%	12%	14%
Neutral	38%	70%	75%	73%
Positive	38%	4%	13%	13%
Total	100%	100%	100%	100%
<b>Case-2</b>				
Frequency Crosstabulation				
Negative	3	8	22	33
Neutral	3	19	130	152
Positive	0	2	8	10
Total	6	29	160	195
Percentage Crosstabulation				
Negative	50%	28%	14%	17%
Neutral	50%	66%	81%	78%
Positive	0%	7%	5%	5%
Total	100%	100%	100%	100%
<b>Case-3</b>				
Frequency Crosstabulation				
Negative	4	5	27	36
Neutral	0	13	120	133
Positive	0	1	19	20
Total	4	19	166	189
Percentage Crosstabulation				
Negative	100%	26%	16%	19%
Neutral	0%	68%	72%	70%
Positive	0%	5%	11%	11%
Total	100%	100%	100%	100%

	<i>Significance level</i>	<i>Missing data</i>
Peter and Søren case	0.028	130
Jeannette and Pernille case	0.06	148
Anders and Jørgen case	0.06	154

In case-1, the odds ratio of finding the relational influence negative to neutral, while also finding outcome distribution fair relative to finding it unfair is 4.12 [ $p > 0.1$ ; 90% CI (0.872, 19.480)]. The odds ratio of finding it negative to neutral while also finding it neutral relative to finding the distribution of the outcome fair is 2.32 [ $p > 0.1$ ; 90% CI (0.965, 5.568)], and the odds ratio of finding it neutral to positive while also holding a neutral opinion on distributive fairness compared with finding it to be fair is 2.82 [ $p > 0.1$ ; 90% CI (0.496, 16.080)]. Since none of results are statistical significant, these may solely be viewed as possible indications of relations. In the first case, then, one is more likely find the relational influence negative to neutral if one also finds the distribution unfair compared to if one found it fair, just as one is more likely to find it to represent a neutral relational value to a positive one if one also holds a neutral opinion on whether it is distributive fair compared with finding it fair. Regardless of this, the majority of the respondents who found the outcome distribution fair also found the relational value to be neutral. Hence, it might be speculated that a vertically fair distribution represents a neutral relational influence or value.

In case-2, the odds ratio of finding it to represent a negative to a neutral relational influence while also finding the outcome distribution to be fair relative to a neutral position on the outcome distribution is 2.49 [ $p < 0.1$ ; 90% CI (1.126, 5.497)] and the odds ratio of finding it to represent a neutral to a positive relational influence while also holding a neutral position relative to finding the distribution of the outcome fair is 0.58 [ $p > 0.1$ ; 90% CI (0.149, 2.291)]. From this it seems to follow that one is more likely to find the relational negative if one also holds a neutral opinion on the fairness matter relative to finding it fair, just as one is less likely to find the relational value neutral to positive if one also holds a neutral opinion compared to finding it fair or equivalent one is more likely to find the relational influence neutral if one also finds the outcome distribution vertically fair.

In case-3, the odds ratio of finding the relational influence negative to neutral while also holding a position neutral relative to finding the outcome distribution fair is 1.71 [ $p > 0.1$ ; 90% CI (0.670, 4.362)], while the odds ratio of finding the relational influence to be neutral to positive and also holding a neutral position relative to finding it fair is 2.06 [ $p > 0.1$ ; 90% CI (0.354, 11.965)]. Although none of the relations are signif-

icant, it seems that a particular trend can be identified. It seems that one is slightly more likely to find the relational influence negative to neutral if one also holds the neutral position relative finding the distribution of the outcome fair or equivalently, one is more likely to find the relational influence neutral to negative if one finds the distribution of the outcome fair.

Across the three cases, a similar pattern can be identified; finding the distribution fair increases the chance of finding it to represent a neutral relational influence. In that sense, if the outcome distribution is found to be fair, then it is also found to have no relational influence. Although, across case comparison is problematic, the consistent pattern found here across the cases seems to suggest, although not statistically significant, that a fair outcome distribution will have influence on the relation between the receiving and non-receiving individual. From a third-party perspective, then, Lawler's (2003) assumption seems to be correct, just as it seems to support the Aristotelian perspective that the fair is the middle, chapter II and V. Though, relational concerns might influence first-party judgment, from a third-party perspective, the fair will have no relational influence. This should, of course, also be viewed relative to the fact that the respondent did not find that a norm violation had occurred in neither of the cases. As a consequence of this, it is, as such, problematic to speculate on the possible influence of norms on this relation.

*Result-3a:* Across all the cases, then, a similar pattern can be identified. Finding the outcome distribution to be fair increases the chance of also finding it to represent a neutral relational influence relative to finding it to represent a positive or negative relational influence.

Figure 52

**Frequency and Percentage Crosstabulations of the Relation between Relational Influence and Norm Violation**

Relational Influence	Norm violation			Total
	No norm violation	Neutral	Norm violation	
<b>Case-1</b>				
Frequency Crosstabulation				
Negative	22	5	3	30
Neutral	133	21	1	155
Positive	23	0	5	28
Total	178	26	9	213
Percentage Crosstabulation				
Negative	12%	19%	33%	14%
Neutral	75%	81%	11%	73%
Positive	13%	0%	56%	13%
Total	100%	100%	100%	100%
<b>Case-2</b>				
Frequency Crosstabulation				
Negative	31	0	2	33
Neutral	149	3	1	153
Positive	8	2	0	10
Total	188	5	3	196
Percentage Crosstabulation				
Negative	16%	0%	67%	17%
Neutral	79%	60%	33%	78%
Positive	4%	40%	0%	5%
Total	100%	100%	100%	100%
<b>Case-3</b>				
Frequency Crosstabulation				
Negative	27	8	1	36
Neutral	122	12	0	134
Positive	18	2	0	20
Total	167	22	1	190
Percentage Crosstabulation				
Negative	16%	36%	100%	19%
Neutral	73%	55%	0%	71%
Positive	11%	9%	0%	11%
Total	100%	100%	100%	100%

	Significance level	Missing data
Peter and Søren case	0.00	130
Jeannette and Pernille case	0.014	147
Anders and Jørgen case	0.042	153

From the relation between relational influence and fairness, I turn to the relation between relational influence and norm violation. In case-1, the odds ratio of finding the relational influence negative to neutral, while also finding that no norm violation has occurred relative to the neutral position is 0.69 [p > 0.1; 90% CI (0.281, 1.717)]. Because no statistically significant relation can be established, one can solely indicate that

a possible relation could be that one is less likely to find the relational influence negative to neutral if one also finds that no norm violation had occurred compared to holding a neutral position.

A similar pattern can be identified in the two other cases. In case-2, the odds ratio of finding the relational influence to be neutral to positive while also finding that no norm violation has occurred relative to holding a neutral position is 12.42 [ $p < 0.1$ ; 90% CI (2.455, 62.799)] indicating that one is far more likely to find the relational influence neutral to positive while finding that no norm violation had occurred relative to holding a neutral opinion. In case-3, the odds ratio of finding the relational influence to be negative to neutral while finding that no norm violation has occurred relative to holding a neutral position is 0.33 [ $p < 0.1$ ; 90% CI (0.145, 0.762)] indicating that one is less likely to find the relational influence negative to neutral if one also finds that no norm violation had occurred compared to if one held a neutral opinion – or equivalently, one is more likely to find the relational influence negative to neutral if one held a neutral opinion.

*Result 3-b:* Finding that no horizontal norm violation have occurred relative to holding a neutral opinion, increases the likelihood of finding the relational influence neutral, just as finding that no norm violation has occurred increases the likelihood of finding the relational influence neutral to negative or positive.

I now turn from the relation between horizontal fairness and relational influence to the relation between the emotional value of not receiving and relational influence. In the correlation matrix in figure 46 it can be seen that the emotions of receiving are low correlated with relational influence across all the cases, whereas the emotions of not receiving are fairly positive correlated with relational influence, indicating that when emotions of not-receiving goes from negative to negative to positive, relational influence also goes from negative towards positive. Moreover, this might also be interpreted as an indication of the source of this relational influence, insofar as the results from the correlation matrix seems to suggest that relational influence might be more related with the emotional experience of the non-receiving agent than with the emotional experience of the receiving agent. That is, the relational influence or dynamics may stem from the

emotional experience of the non-receiving agent. The relational influence, then, primarily expresses the emotional experience of the non-receiving agent. It should, of course, be noted that even though the correlation is statistically significant it cannot be interpreted as more than an indication due to the rather low positive correlation.

Figure 53

**Frequency and Percentage Crosstabulations of the Relation between Relational Influence and Emotions of Not Receiving**

Relational Influence	Emotions, not receiving			
	Negative	Neutral	Positive	Total
<b>Case-1</b>				
Frequency Crosstabulation				
Negative	21	5	4	30
Neutral	29	110	14	153
Positive	10	7	11	28
Total	60	122	29	211
Percentage Crosstabulation				
Negative	35%	4%	14%	14%
Neutral	48%	90%	48%	73%
Positive	17%	6%	38%	13%
Total	100%	100%	100%	100%
<b>Case-2</b>				
Frequency Crosstabulation				
Negative	18	12	3	33
Neutral	23	121	9	153
Positive	0	5	5	10
Total	41	138	17	196
Percentage Crosstabulation				
Negative	44%	9%	18%	17%
Neutral	56%	88%	53%	78%
Positive	0%	4%	29%	5%
Total	100%	100%	100%	100%
<b>Case-3</b>				
Frequency Crosstabulation				
Negative	28	6	2	36
Neutral	27	95	12	134
Positive	4	9	7	20
Total	59	110	21	190
Percentage Crosstabulation				
Negative	47%	5%	10%	19%
Neutral	46%	86%	57%	71%
Positive	7%	8%	33%	11%
Total	100%	100%	100%	100%

	Significance level	Missing data
Peter and Søren case	0.00	132
Jeannette and Pernille case	0.00	147
Anders and Jørgen case	0.00	153

From the correlation matrix, I turn to an analysis of the relation between relational influence and the emotions of not- receiving. In case-1, the odds ratio of finding the relational influence to be negative to neutral while finding the compensatory difference to represent a negative emotional value relative to finding it to represent a neutral emotional value is 15.93 [p < 0.1; 90% CI (6.540, 38.809)], the odds ratio of finding the relational influence negative to positive while finding the compensatory difference to represent a neutral emotional value to the non-receiving agent relative to finding it to represent a positive emotional value is 0.16 [p < 0.1; 90% CI (0.048, 0.529)]. Hence, the likelihood of finding the relational influence to be negative to neutral while also finding the compensatory difference to represent a negative emotional value is greater than if one found it to represent a neutral emotional value, just as one is less likely to find the relational influence to be negative to neutral if one finds it to represent a neutral emotional value compared to if one found it to represent a positive emotional value.

A similar tendency can be identified in the other two cases. In case-2, the odds ratio of finding the relational influence negative to neutral while also finding the compensatory difference to represent a negative emotional value to the non-receiving individual relative to finding it to represent a neutral emotional value is 7.89 [p < 0.1; 90% CI (3.840, 16.219)], whereas the odds ratio of finding the relational influence negative to neutral while finding the emotional value neutral relative positive is 0.30 [p < 0.1; 90% CI (0.089, 0.996)]. This seems to indicate that one is more likely to find the relational influence negative to neutral if one also finds the emotional value to the non-receiving to be negative compared to if one found it to be neutral, just as one is less likely to find it negative to positive if one found the emotional value neutral compared to if one found it positive. In case-3, the odds ratio of finding the relational influence negative to neutral while also finding the emotional value negative relative to finding it neutral is 16.42 [p < 0.1; 90% CI (7.196, 37.464)], whereas the odds ratio of finding the relational influence negative to neutral while also finding the emotional value neutral relative to finding it positive is 0.38 [p > 0.1; 90% CI (0.090, 1.598)].

*Result-3c:* The likelihood of finding the compensatory difference to represent a neutral relational influence increases if it is also found to represent a neutral emotional value to the non-receiving agent relative to finding it to represent a negative or positive emotional value.

Figure 54

**Frequency and Percentage Crosstabulations of the Relation between Relational Influence and Emotions of Receiving**

Relational Influence	Emotions, receiving			
	Negative	Neutral	Positive	Total
<b>Case-1</b>				
Frequency Crosstabulation				
Negative	0	2	28	30
Neutral	2	33	120	155
Positive	1	1	26	28
Total	3	36	174	213
Percentage Crosstabulation				
Negative	0%	6%	16%	14%
Neutral	67%	92%	69%	73%
Positive	33%	3%	15%	13%
Total	100%	100%	100%	100%
<b>Case-2</b>				
Frequency Crosstabulation				
Negative	0	7	26	33
Neutral	0	12	140	152
Positive	0	0	10	10
Total	0	19	176	195
Percentage Crosstabulation				
Negative	0%	37%	15%	17%
Neutral	0%	63%	80%	78%
Positive	0%	0%	6%	5%
Total	100%	100%	100%	100%

	Significance level	Missing data
Peter and Søren case	0.037	130
Jeannette and Pernille case	0.069	148

Above, I analyzed the relation between relational influence and the emotions of not receiving. Now, I turn to the relation between relational influence and the emotions of receiving. The relation in case-3 is not significant; hence I focus on the two other cases. In case-1, the odds ratio of finding the relational influence to be negative to neutral while finding the reward to represent a neutral emotional value to the receiving individual relative to finding it to represent a positive emotional value is 0.26 [ $p < 0.1$ ; 90% CI (0.074, 0.907)], and the odds ratio of finding the relational influence to be neutral to positive while also finding the emotional value to be neutral relative to finding it positive is 7.15 [ $p < 0.1$ ; 90% CI (1.290, 39.628)]. This seems to indicate that one is less likely to find the relational influence negative to neutral if one finds the emotional value

to be neutral compared to finding the emotional value positive or equivalently one is more likely to find the relational influence negative to neutral if one also finds the emotional value to be positive, just as one is more likely to find the relational influence to be neutral than positive if one also finds the reward to represent a neutral emotional value to the receiving agent compared to finding it to be positive. In case-2, the odds ratio of finding the relational influence to be negative to neutral while also finding the emotional value to be neutral compared to finding it positive is 3.14 [ $p < 0.1$ ; 90% CI (1.329, 7.424)] indicating that one is more likely to find the relational influence to be negative to neutral if one finds the emotional value to be neutral than if one found it to be positive or equivalently one is less likely to find the relational value negative to neutral if one also finds the emotional value positive compared to finding it neutral.

*Result-3d:* The likelihood of finding the relational influence to be neutral while also finding the reward to represent a positive emotional value to the receiving individual is greater than if the emotional value was found to be neutral.

From this it seems to follow that while the respondents find the compensatory difference to represent a negative emotional value to the non-receiver and the reward to represent a positive emotional value to the receiver, they do not find it to influence the relation between the receiver and the non-receiver. Although, the number of cases prevents an isolation of the individual effects of fairness and norm violations, insofar as the majority of the respondents found the cases included in the survey to be fair and judged that no norm violation had occurred. However, it might be argued that if the compensatory difference is found fair and no horizontal norm violation has occurred, then, the negative emotions of the non-receiving individual and the positive emotions of the receiving individual will have no real influence on the relation. In that sense, it might be speculated that the effect of the emotions are cancelled out by the judgments related with vertical fairness and horizontal norms.

A compensatory difference may, then, represent a negative emotional value to the non-receiving individual, just as the underlying reward may represent a positive value to the receiving individual. However, the positive emotions of the receiving individual

may not be viewed as the driver of positive relational influence, insofar as relational influence was not correlated with this variable; rather it may be argued that insofar as the relational influence is positively correlated with the emotions of the non-receiving agent, what drives the relational influence towards neutral is the emotions of the non-receiving agent. The question, in other words, might be why the negative experience is not represented as a negative relational influence. In relation to this, it might be argued that negative emotions associated with a compensatory difference will not necessarily be directed towards the other, the non-receiving agents may as a consequence internalize the negative externality generated by the actions of the other – implying that these will have no real relational effect.

The results produced here, at least to some extent, contradicts the assumptions in equity theory and especially Fehr and Schmidt (1999), because here equality would be associated with neutrality, although as argued by Aristotle, the equal is not necessarily the fair, insofar as the fair is that which maintains some status quo. What seems interesting here is, however, that the fair is not that which maintains some relational status quo; rather it is that which reflects the *new* relational status between the parties, this might of course be due to the third-person perspective. In other words, finding a compensatory difference fair and being the result of non-norm violating actions may lead to acceptance, but it may not prevent the non-receiving agent from associating the difference with negative emotions. A social norm may exist stating that if a compensatory difference is fair and caused by fair actions and principles, then the negative emotions invoked by this should not be transformed into a negative relational influence.

## **10.6 Tentative Conclusions**

The aim of this chapter has been to discuss some of the theoretical points developed in the previous chapters against a small empirical analysis based on three vignettes. In juxtaposition to the perspective discussed so far in this thesis, the perspective shifts from a first-person perspective to a third-person perspective. This change of perspective is associated with the introduction of the vignettes, as argued in chapter IX. This shift of perspective also allows a study of the generative mechanisms, insofar as it allows one to collect data at a more abstract level – the patterns of judgment and the cognitive struc-

tures on which these cognitive structures are based, combined with the theories previously discussed, see chapters I and IX.

The present chapter has been divided into two main parts, one focusing on identifying the general trends in the data set and another part focusing on analyzing the relations between the different dependent variables. The independent variables can, here, be viewed as being the items on which the vignettes are based – the descriptions of the three small histories which produces the context of the respondents' judgment in combination with their sense making – what they read into these, the information they retrieve in order to produce a particular judgment and the cognitive structuring which mediates the information and the actual judgment, see chapter IX.

In the first part of this chapter, the general trends in the data set were identified. It was identified that the majority of the respondents across all the cases found the after-situation to be distributive as well as procedural fair. That is, the after-situations in the three cases were found to be characterized by a fair distribution, just as the foundation of distribution was found to be fair. Hence, the after-situations across the three cases can be viewed as being vertically fair. In case-1, the after-situation was brought about by the receiving agent who after having participated in a project team obtained a bonus. The receiving agent initially obtained an outcome below that of the non-receiving agent, but after obtaining this bonus they obtained an equal outcome. In case.2, the after-situation was brought about by the receiving agent's choice of participating in an on-the-job course which earned her a lasting qualification bonus. Initially, the receiving agent obtained an outcome below that of the non-receiving agent, but in the after-situation her outcome was above that of the non-receiving agent. In case-3, the after-situation was not brought about by the receiving agent's deliberate differentiation; rather he was offered the opportunity of obtaining an MBA-degree because he showed interest in management related tasks. From a vertical perspective, the receiving agents in all the cases did something to obtain their reward – they showed agency, in the sense that they did something which indirectly brought the after-situation about. In one the interviews the interviewee emphasized this as one of the reasons she had for finding it vertically fair. By showing agency, the receiving agent earned her extra outcome. In juxtaposition to the theories discussed in chapters V and VII, the fairness of the after-situation is not

assessed based on a comparison between the before and the after-situation, rather it is based on something which lies between these two extremes of the relational and temporal dimensions of the constitution of the fair outcome claim. To some extent, then, it might be argued that the vertical fairness of some outcome is also contingent upon the actions performed by the receiving agent and not just a comparison. Besides this focus on the actions of the receiving agent, the foundation of after-situation needs to be acceptable to the non-receiving agent – that is, the non-receiving agent needs to be informed about the foundation of the distribution, the reason is not so much a matter of an assessment of whether it is in accordance with the agent's own values, rather it seems to be related with whether or not the after-situation was directly caused in an acceptable manner – based on fair principles. This particular understanding of procedural fairness is similar to the perspective discussed by Lawler in chapter II, although he emphasized the rational judgment dimension of this. From a somewhat different perspective, however, this perspective seems to be in juxtaposition to the relational perspective on procedural fairness emphasizing the socioemotional structuring of the vertical relation, see chapter III. This socioemotional nature of the vertical relation is, nevertheless, present in the symbolic dimension of outcome differences, because as argued by one of the interviewees, what matters is not so much the material outcome difference between the before and the after, rather it is the symbolic difference which matters, in the sense that when the after-situation deviates from the before-situation, the difference matters because it is transformed into a difference in vertical appreciation. On the other hand, the difference does not matter horizontally and if it does it might be related with the horizontal relation itself. In other words, vertical fairness judgments seems here to be more related with the causal process of the reward situation than with comparisons between a before and an after.

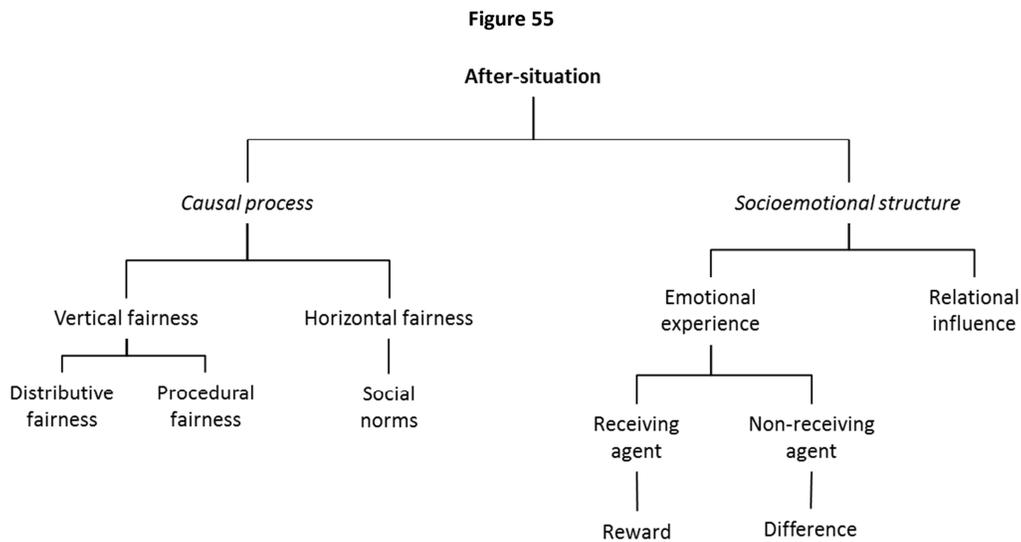
This focus on the causal processes of the reward situation leads, in accordance with the discussion in chapter VI, to a distinction between the direct cause of the after-situation and the indirect cause. While the former may be associated with the actions of management and, as such, related with the vertical relation and vertical fairness, the latter may be associated with the horizontal relations and the behavior of the receiving and the none-receiving agents. In accordance with Folger and Cropanzano's perspective, one of the interviewees argues that management is the direct cause of the after-

situation and that the actions of employees may not change much in terms of the after-situation. Across all the cases, the respondents found that none of the receiving agents had violated some horizontal unwritten social rule by actively differentiating themselves. As argued by one of the interviewees, there are not, as such, any social rules preventing the agents from showing active agency and trying to bring some state about unless this action might be socially inappropriate – that is, the employees are allowed to be egoistic and ambitious, but only as long as they do not hurt their colleagues, because as the interviewee argues, no one likes to be disliked and one has to make it work. Thus, what is here referred to as appropriate behavior can be resembled with what I have called social rationality and the aim of this principle of action is, on the one hand, to maintain the social relations; and, on the other hand, the aim is also to satisfy the agent's psychological need to belong. In addition to what was argued in chapter VI, a third kind of causal agency may be taken into consideration, insofar as the actions of the non-receiving agent may also influence the after-situation. In case-2, the non-receiving agent is offered a similar opportunity as the receiving agent, but declines this and as argued by one of the interviewees, this agent has only herself to blame. From a causal perspective, then, one needs to distinguish between the indirect causal relation between the actions of the receiving agent and the after-situation, the direct causal agency of management, and the direct causal agency of the non-receiving agent.

In chapter VII, it was argued that if the after-situation deviates from the before-situation, this deviation might, from a first-person perspective, be hedonically experienced as a loss. From a relational perspective the loss may be associated with a loss of status and from a temporal perspective manifesting itself psychophysically as hedonic experience and, as such, it might be associated with a loss of satisfaction or motivation. A loss should, in other words, manifest itself psychophysically as a painful or negative emotional experience whereas a gain should manifest itself as a pleasant or positive emotional experience. If this gain, however, happens at the non-receiving agent's expense, then, one might expect this to influence the relation between the receiving and the non-receiving agent negatively – that is, in juxtaposition to the Fehr-Schmidt model on quasi-moral norms in which such a gain at the other's expense would be associated with guilt or Elster's account in which it would be associated with shame, in the present study it is simply analyzed as a loss of belongingness measured qualitatively as changes

in the foundational emotional structuring of the relation. This focus enables one to push past conceptual disagreements on the nature of social norms and their emotional resonance, see chapter VIII. From this line of reasoning, it also seems to follow that the psychophysical experience of fairness in vertical as well as horizontal relations lies between pleasure and pain, as also argued by Aristotle, as the neutral experience – that which does not cause a negative relational influence. That is, what is argued here is not that the non-receiving will not attach negative emotional experience to not receiving or that the receiving agent will not attach positive emotional experience to receiving, rather what is argued here is simply that these individual emotions will not have any relational influence. The relational influence cannot be viewed as a result of opposite emotions in the sense that if the negative emotional experience of the non-receiving agent exceeds the positive emotional experience of the receiving agent, the relational influence will be negative. To understand why this cannot be so and the most likely relational influence might be neutral, one has to take horizontal and vertical fairness into account. Because even though the majority of the respondents across the three cases found the emotional experience of the non-receiving agents to be either negative or neutral and the emotional value of receiving to be positive, the relational influence was primarily found to be neutral. A vertically fair outcome distribution is more likely to be associated with a neutral emotional experience from the non-receiving agent's perspective, just as a horizontally fair action performed by the receiving agent is also found to be associated with a neutral emotional experience. On the other hand, from the perspective of the receiving agent, the horizontally and vertically fair reward is associated with a positive emotional experience. Furthermore, the horizontally and vertically fair compensatory difference is also more likely to be associated with a neutral relational influence. In such cases, then, when the after-situation is perceived as being fair, the non-receiving who might experience negative emotions is more likely to internalize the effect of this negative externality, in the sense that she, as argued by one of the interviewees, does not just blame those around her even if she might feel bad. This might follow as a consequence of the principle of social rationality which seems to govern horizontal social interaction. In some situations, the emotional experience of the non-receiving agent may, of course, influence the relation and in these cases, it seems that, negative emotional experience is associated with a negative relational influence – that is, the variables are positively corre-

lated. Although, the correlation between relational influence and the emotional experience of the receiving agent was not statistical significant, the two variables are negatively correlated meaning that as the emotional experience of the receiving agent change from positive to negative the relational influence changes from negative to positive. However, the two agents' emotional experience cannot be viewed as producing a simple equation from which relational influence can be determined as the residual, insofar as the fairness of after-situation needs to be taken into account. In accordance with this, one might perhaps distinguish between two different dimensions of the after-situation: (1) a dimension related with how it was brought about; and (2) a dimension related with the socioemotional structuring of this associated with the emotional experience of the receiving and non-receiving agents, and the relational influence. The following relation can now be illustrated:



From this framework it seems to follow that the after-situation is brought about in a particular manner and this causal process is contingent upon how the distribution is assessed, just as the foundation of this distribution. Matters related with the distribution and its foundation can be associated with vertical fairness and the direct causality of the reward process. The situation is also brought about by the actions of receiving and non-receiving, in the sense that their causal agency might directly/indirectly cause the after-situation. In juxtaposition to vertical fairness which might either be related with out-

come or the procedures, horizontal fairness is primarily related with actions and not outcomes. Psychophysically, these matters associated with horizontal and vertical fairness manifests themselves as the individual agents' emotional experience as well as relational influence. Thus, the experience of the situation cannot, as argued in chapter VII, be reduced to matters solely related with comparisons between before and after or fairness.

## Conclusion

The primary research question on which this work is centered concerns an exploration of how the fair claim is constituted in, on the one hand, a compensation system based on the two values individualization and individuality; and, on the other hand, a normative relational social system. Based on the exploration of these two systems, a third question follows: how might potential deviations influence the horizontal dimension?

Although, the aim of this thesis is to discuss the fair claim, the aim is not to tap into the discussion between equity and equality theorists on which principle is fairest, rather the aim is to discuss the constitution of the claim – treating the fair claim as a case in point or a *Sachverhalt* enables me to push past these discussions and tap into how the fair claim is constituted, but also the content of the claim in the sense that the fair claim cannot be reduced to evaluations of outcomes or, at least, not in a material sense, because the fair claim also encompasses an interaction-related and non-consequentialistic dimension related with social agents acts one another. That is, the fair claim is not solely a matter of being treated fairly with regard to obtaining a fair share, insofar as it also encompasses an interactional dimension. This focus on the constitution of the claim also follows from the philosophical foundation of thesis, in the sense that the aim is not so much to study a claim as it manifests itself empirically in, for example, wage negotiations, as it is to understand how it is produced – why it is held as a defensible fair claim. In that sense, the focus shifts from the empirical manifestation of the claim to the constitution of this or the generative mechanisms on which this is based. While the claim is held by the individual, the causal process which led her hold this claim as defensible cannot, necessarily at least, be reduced to an internal causal process, inasmuch as this claim may be the product of several causal chains interconnecting with one another stemming from different analytical levels. Even though, the claim may not be viewed as being the social agent's own constructing, this should not be interpreted as arguing that the individual does not have any reasons to hold it.

Additionally, the aim of this thesis is not to tap into the discussion between expectancy theorists and fairness theorists on the matter of motivation – that is, whether social agents are motivated by monetary outcome or fairness. From the point of view devel-

oped in this thesis, the question is not whether individuals are extrinsically or intrinsically motivated or whether they are motivated by a quasi-moral desire to be or at least appear to be fair, insofar as no easy answers can be produced on this matter and the question cannot be reduced to an either/or question, because individuals are motivated by their monetary outcome at least in the sense that they perceive it as an means to an end, as argued by Lawler (2003) in chapter II, but this instrumental explanation may not constitute the entire explanation behind human motivation, insofar as Deci and Ryan (2000) argues that humans are driven from the inside by an intrinsic will to achieve and to realize their true self and in continuation of this line of reasoning, Frey (1997) argues that monetary compensation may out-crowd this internal sense and public service motivation theorists, such as Andersen (2012), argues that in addition to this potential out-crowding effect following from non-supporting control, one also has to take into account that employees may be motivated by a desire to make a difference – an altruistic desire (see chapters IV and VIII). From a somewhat different perspective, the early equity theorists, such as Jaques (1961) or Adams (1963), argues that social agents are motivated by a desire to reduce cognitive dissonance and this desire leads them to match their input with the outcome they obtain and, as such, these social agents are driven by an equitable sense of fairness (see chapter V). Modern equity theorists, such as Fehr and his colleagues, recognize this foundational pattern of behavior, but argue that this is the enactment of the quasi-moral principle of reciprocity (see chapter IV).

From this it seems to follow that human motivation, in a vertical perspective, cannot be reduced to either/or explanations, but must allow for different explanations to be connected with one another. Nevertheless, the main approach to this problem in the present thesis is more pragmatic in nature, in the sense that it is assumed that social agents are motivated by both extrinsic and intrinsic matters, just as they are motivated by egoism, envy, etc. and non-egoistic motives such as reciprocity, guilt, shame, etc. (see chapters III, IV, V and VIII). Additionally, it is also assumed that social agents act instrumentally as well as non-instrumentally rational (see chapter III). With regard to the different motives and motivations which might drive social agents, the question is how they balance all these different drivers, rather than which of them might dominate.

In chapter II, I discuss the nature of the fair claim in contemporary theories on compensation management with a theoretical departure in Lawler's (1990; 2000; 2003) comprehensive theory and Schuster and Zingheim's (1992; 2000) approach which can, at least to some extent, be viewed as parallel to Lawler's project. The fair claim discussed by these authors is vertical in nature, in the sense that it primarily concerns issues related with fairness in the organization-employee relation. The fair claim can, however, not be reduced solely to matters related with what characterizes a fair treatment, insofar as it is more extensive in nature and directed at the employment relationship in general. That is, in juxtaposition to compensation systems based on, for example, job analysis, the aim is here not *simply* to secure horizontal and vertical fairness along the lines of the hierarchy or an impartial evaluation solely based on the job, rather the aim to personalize the employment relation – to recognize the individual who performs the job and to recognize her as an unique and rational individual who should be allowed to let her individuality in the form of choices be reflected in the rewards she obtains. This recognition of individuality should, for example, be reflected in the rewards obtained by the individual, in the sense that she should be allowed to design her compensation mix in accordance with her interests as a rational individual; hence this can also be viewed as a reduction of the paternalistic flavor accompanying fixed reward packages designed by management. Furthermore, the compensation system should be based on egalitarian values, in the sense that rewards should not be contingent on jobs held or hierarchical positions, rather all members of the organization should be offered similar choices, just as rewards should solely be contingent on the individuals' choices – in order to secure a line of sight, implying that the employee should perceive the reward as within her control to bring about and, not at least, as feasible. These choice contingent rewards, redirects the focus of the compensation system from matters related with job evaluation to matters related with evaluation of individual performance and competences. Compensation can, as a consequence of this, not be equated with pay, because it encompasses several dimensions, thus total compensation emerged. The base-pay should be based on the individuals' competences, the variable component on some kind of performance related assessment, and the benefits on a self-composed reward package.

The theoretical shift from job-based to competence-based organizations followed from a theoretical shift in the understanding of the organization's core, but also from the recognition that individuals change organizations, in the sense that the core foundation of the organization is not jobs, but the individuals who perform these and in order to develop the organization, one must develop the employees. Hence, humans became the core of the organizations and in particular their competencies. With this also followed that the employees should be willing to develop their competences in accordance with the organization's needs to secure a more dynamic foundation. This focus on continuous development can also be viewed as an extension of the negotiation on which the employment relationship is based, in the sense that besides being a continuous negotiation of the distribution of tasks, within some zone of acceptance, it now also encompasses education and personal development – matters which are not, necessarily at least, governed by the same zone of acceptance as governs task-related negotiations.

The recognition of individual choice should also be related with the individual's performance – qualitatively as well as quantitatively. In cases based on individual performance, performance differences may lead to differences in outcome related with the individual's performance within a given period evaluated against some predetermined goals. Hence, in monetary terms, two employees holding similar jobs within the same organization may obtain different amounts of monetary compensation, because it is contingent upon an assessment of their competences as well as their performance, just as their compensation may also be qualitatively different due to different choices related with the composition of the reward package. From this focus on total compensation, then, follows that compensatory differences cannot solely be reduced to quantitative matters. Moreover, from this also follows that horizontal comparison may be problematized by this addition of the qualitative matter, insofar as this requires comparison of the satisfaction which the individual obtains from these self-composed reward packages. Lawler's (2003) proposed solution to these qualitative fairness issues is simply to tell the employees that these reward packages are self-composed.

From this focus on the individual and the recognition of individuality follows also a high-degree of empowerment, but also responsibility in the sense that the employee becomes responsible for her own outcome – for designing her reward package and de-

livering the expected performance. Moreover, from this focus also follows that compensatory differences cannot be viewed as emerging from the different jobs held and, as such, the concept of organizational classes discussed by Baron and Pfeffer (1994) loses some of its explanatory power, because compensatory differences are not as such imposed from the outside, rather they emerge from the inside, as actions on differentiation performed by the employees. Quantitatively, the employees may differentiate themselves from one another by being better or working harder and qualitatively they may differentiate themselves based on different choices regarding the composition of the reward package, but also the educational opportunities they pursue. Of course, jobs as well as hierarchical positions still exist, but they lose some of their explanatory power, insofar as horizontal differences may emerge from the outside – exogenous differentiation – but it can also emerge from the inside – endogenous differentiation. Recognition of individuality is, then, accompanied by an assumed will to differentiation – to be different and to be recognized as such. The fair claim, in other words, is constituted on the two foundational values – individualization and individuality manifested as compensatory differences and acts on differentiation. The fair claim is a claim on being recognized as a unique and rational individual who have the right to control her own processes of choice and being recognized for the choices she makes. The vertical claim emerges from choices and matters related with competency and performance level. The vertical claim is horizontally defensible if: on the one hand, it is based on individual choices reflecting the interests of the individual; and, on the other hand, if the compensation is properly determined, that is, it is horizontally fair if it is in accordance with pre-established principles explained in the organization's compensation policy and if the measures applied are accurate. Horizontal fairness, then, is reduced to a matter of rational acceptance – based on the acceptance of facts.

Firstly, it may be questioned whether the employment relation can be reduced to such a quasi-libertarian system without losing some of its moral foundation; secondly, it may be questioned whether horizontal fairness can be reduced solely to matters related with rational acceptance; and thirdly, it may be questioned whether acts of differentiation might not undermine the motivational base of the compensation system if this destroys social bonds between the agents which might lead to a loss of belongingness which, subsequently, might lead to a loss of motivation to achieve.

On the question whether the employment relation can be reduced to quasi-libertarian system without losing some of its moral foundation, I argue in chapter IV that the employment relation may be viewed as being partly based on the quasi-moral norm reciprocity and partly based on authority. In accordance with the perspective developed by Fehr and his colleagues, it might be argued that the employment relation is centered upon normative expectations related with reciprocity. In accordance with Rabin (1993) and Aristotle (350BC/2007), reciprocity may be defined as acting towards the other in a like manner as she acted towards oneself, but reciprocity does not solely encompass actions it is also related with believed intentions, in the sense that the evaluation of the action is not solely consequentialistic and based on an evaluation of consequences, but also takes into consideration the non-instrumental dimension related with whether the individual performed the action with a particular intention in mind. In that sense, an action is evaluated on whether its intention is in accordance with governing relational norms. From this perspective, an action framed in a particular way changes the underlying nature of the game, in the sense that it changes the perceived value attached to different opportunities and, as such, by invoking different norm sets, different patterns of interaction emerge. An employer, for example, who offers a contract with a contingent bonus may invoke a set of norms associated with economic interactions, whereas if the employer had offered a non-contingent employment contract (a trust-contract) with a pay perceived as generous by the employee, he might have invoked a different norm set based on reciprocity. This activation of different norm sets may be called belief-management, in the sense that the employer can, through his actions, signal different kinds of intentions which might lead to different kinds of beliefs regarding future actions – that is, the employer cannot only influence how he is perceived by the employee, but also how the employee's beliefs about how the employer perceives her. For example, by offering a generous trust contract, the employer is simultaneously signaling his own kind, just as he is signaling that he perceives the employee as a reciprocal player who will reciprocate his generous offer. Whether this strategy is successful is, however, contingent upon the nature of employee – whether she holds the value reciprocity – just as it is contingent upon the behavior of other players, thus even an employee holding the value reciprocity might not reciprocate if she believes that this value is not held by her fellow employees.

In that sense, different kinds of contracts may signal different kinds of intentions and as a consequence, influence the norms which come to govern the future normative foundation of the vertical relation. Empirical studies also show that it is not just different contracts which might have this influence, insofar as just mentioning money may prime the agents' mindset in a particular manner and invoke different norm sets, and when a norm set has been invoked it binds – so, a previously generous employer who suddenly changes the organization's pay policy to one based on bonuses may destroy the quasi-moral obligation accompanying the principle of reciprocity and reduce future interaction to matters solely related with economic norms. The employer may not, however, be perceived as unfair, because by changing the nature of the game, he also changes how fairness is defined. This can be viewed in opposition to Kahneman, Knetsch, and Thaler (1986a; 1986b) who argues that the fair is that which does not deviate from the past – in the sense that a negative deviation between past and present is regarded as a loss.

The employment relation, however, cannot be reduced to matters solely related with moral norms, because it naturally implies asymmetry of power and, as such, authority. From this authority relation follows that the employer is allowed to, within the zone of acceptance, decide different matters related with the job. Of course, due to the rather particular nature of the employment contract, this should not be perceived in a static manner, insofar as dynamics are introduced through continuous negotiation of, for example, the distribution of tasks. Hence, existing simultaneously with the different relational norm sets is also a formalized claim which emerges from the very nature of the employment relation.

From this it seems to follow that the fair claim is contingent on the relational normative foundation governing vertical relations. A fair claim following from a pure economic normative foundation may be significantly different from the one emerging from a relational normative foundation based on reciprocity. While Lawler's approach cannot be reduced solely to a kind of economic rationality, its quasi-libertarian foundation points towards a conception of the vertical relation as not based on any kind of moral obligations binding the economic agents to a particular behavior. In that sense, the employee is solely responsible for herself and her own choices – in a sense this lack

of other-directed responsibility generates a conceptualization of an agent whose actions are solely reflecting her interests. While this sense of non-commitment and lack of other-regarding responsibilities may not be in juxtaposition to the principles of rationality, it represents, at least, a simplification of rationality by liberating it from its moral foundation.

The relation between rationality and morality leads to the second question regarding whether horizontal fairness can be reduced to matters solely related with rational acceptance. Rationality is perhaps one of the most influential ideas in social science, because just as many theories have emerged in accordance with the principles, just as many theories have also emerged in opposition to them. In that sense, from a metatheoretical perspective, rationality can be viewed as a foundational concept permeating social science. The theories on human motivation, discussed above, can be also be categorized in accordance with different principles of rationality, in the sense that Deci and Ryan's perspective as well as Frey's and Andersen's can be viewed as belonging to non-instrumental tradition in rational philosophy associated with the philosophy of Plato (360BC/1961; 380BC/1961) and Kant (1781/1787/2007; 1785/2002; 1788/2004; 1793/1794/2009); while the rational principles of Lawler's theory are similar to the utilitarian tradition in rational philosophy associated with, among others, the philosophy of Hume (1739/1985). On the other hand, some theories emerges in opposition to rational philosophy, Kahneman, Knetsch, and Thaler's theory on the cognitive bases of fairness emerged in opposition to the utilitarian perspective on economic behavior, just as Kahneman and Tversky's (1979) prospect theory, on which the cognitive bases of fairness is also based, emerged in opposition to the utilitarian conception of choice, and Kahneman's (1999) later hedonic psychology is centered upon the philosophy of Bentham (1780/2007). In general, theories on organizational fairness emerged in opposition to economic rationality – a synthesis of Hobbes' (1651/1985), Hume's, the late Smith's (1776/2003), Bentham's, and Mill's (1863/2001; 1874/2000; 1882/2009) respective utilitarian perspectives. Inspired by Aristotlian ethics, the early theorists on equity argued that social agents cared about acting fair and being fairly treated, insofar as deviations from this would lead to states of psychological dissonance – a notion originally inspired by Heider's (1946) cognitive balance theory and Festinger's (1957/1985) theory. Although, the perspective emerged in opposition to economic rationality, it is in-

spired by rational philosophy, in the sense that the focus is on outcome and in particular how the individual obtains this outcome and, as such, it is related with the notion of outcome associated with, for example, Hobbesian rationality, just as theories on procedural and interactional fairness centered upon the notion of the socioemotional structuring of vertical relations in which unfairness becomes signs on exclusion and fairness signs on inclusion seems to have been inspired by the Humean conception of fairness in which fairness understood as adherence to social norms is related with esteem and self-esteem following from the social emotion, sympathy. Hence, while the theories can be viewed in opposition to economic rationality, they are also inspired by utilitarian rational philosophy. Folger (1998; 2012) introduces in opposition to these utilitarian inspired perspectives on fairness, a perspective inspired by Kantian rational philosophy in order to reintroduce a moral foundation of fairness. From a metatheoretical perspective, then, it might be argued that in order to understand modern theories in social science, one has to understand their intellectual history which is closely linked with rational philosophy. Moreover, it also seems crucial to understand that whereas economic rationality may be liberated from morality, rational philosophy, in general, is not, just as rationality cannot be reduced to economic rationality or instrumental rationality. Besides the metatheoretical perspective, it might also be argued that rational philosophy still holds some key insights into human action and if one is to understand the nature of horizontal relations one needs to understand how humans act in a social world, how they choose their actions in accordance with instrumental reason and morality, just as it might be argued that if horizontal and vertical fairness are to be centered on a similar intellectual heritage, one needs to identify its roots in rational philosophy.

From Platonic rational philosophy emerges a particular kind of rationality, rationality as a way of life. The rational individual should govern her life in a rational manner, in accordance with reason. Rationality does not solely govern actions; it also governs desires, emotions/values, and goals. The Platonic rational agent chooses the action which is most effective in bringing about a true goal based on true desires – desires which are not just truly desired, but also good in a moral sense. Hence, rationality in its Platonic conception is a matter of self-governance in an asocial sense. Hobbes introduces the social level and develops a theory on instrumental rationality in which the rational action is that which is instrumentally rational, within reason, and maintains peaceful

coexistence, in the sense that an action based on unreasonable passions may either drive the individual mad in her pursuit to bring about or it might lead to a state of war – both consequences which are in opposition to the primary principle of Hobbesian rationality – self-preservation. Based on a similar notion of instrumental rationality as Hobbes, Hume argues against the Platonic focus on reason and argues that an action cannot be irrational because it is opposite to reason, because reason is not the opposite desire, just as reason cannot motivate, only desire can motivate. Hence, an irrational action is one which fails to bring about the desired consequences. Moreover, morality as well as social life is governed by hedonic principles related with the concept of utility, in the sense that immoral actions are experienced as negative or painful hedonic experience and moral actions are based on positive or pleasurable hedonic experience. In a similar manner, social life is governed by the sense of sympathy which means that social agents are sensitive to one another, in the sense that their self-perception is influenced by how they believe other social agents perceive them – thus, being badly treated by others influences the individual's self-perception or esteem.

The early Smith (1759/2009) argues against this definition of sympathy and argues that it is because of this sensitivity to others' opinions that social agents are moral, insofar as it allows them to transcend their own immediate situation and take the perspective of the other. Smith also distinguishes between economic rationality related with economic interaction and instrumental rationality and non-instrumental actions either based on an egoistic motive, as the Humean perspective, or real sympathy. More recently, Boudon (2001) has proposed a similar argument not, however, in opposition to the Humean perspective, but in opposition to the Hobbes-Coleman perspective on social life in which matters related with morality are reduced to a rational cost/benefits analysis.

In a perspective somewhat similar to Smith's and in opposition to Humean philosophy, Kant argues that actions springing from a true will must also be in accordance with reason, insofar as the will must be governed by reason. The Kantian rational agent, like the Platonic rational agent, is guided by reason and she chooses her actions as well as goals in accordance with her will. Hence, an action which springs from the will also springs from an intrinsic motivation. Although, Kantian philosophy follows the lines of

Platonic philosophy, Kant's moral philosophy is not solely a matter of self-governance, because it also encompasses principles on how one should treat other individuals, albeit the principles are based on reason so it might be argued that the Platonic asocial agent is simply replaced by an antisocial agent solely governed by her own reason and own will.

From these different positions on the relation between morality and rationality it seems to follow that the rational action is that which is instrumentally optimal, but it is also an action which is in accordance with governing social norms, just as it also an action which is in accordance with the individual's moral principles. An action is, in other words, only rational if it is simultaneously instrumentally rational and socially rational. The social dimension of social rationality follows from the fact that actions always unfold themselves within a social context and in the presence of the other social agents, thus the social agent cannot simply act as if she acted in a social vacuum unmoved by the opinions of others and neither can she act as if her actions did not produce externalities – influenced other individuals. Hence, social rationality emerges as a necessity, because life is social in nature, in the sense that one cannot act in a social context without accepting some minimum of other-directed obligations and expecting that others will do the same. To act, Coleman (1990) argues, is to believe that one is in possession of the right to do so and to interact with others is to believe that one has the right to act in this manner and that one has the right to claim that the other responds in a particular manner. In this perspective, then, social life is based on the exchange of rights and obligations, but in juxtaposition to the Hobbes-Coleman perspective, it might questioned whether this can be reduced solely to a matter of costs and benefits, because as some studies in neuropsychology demonstrates, humans has the capacity to feel empathy, just as they also have a need to belong and, such as, acting against the interests of others may also be to act against one's own interests.

Furthermore, to be rational also implies that one has reasons to act, believe, etc., just as one has reasons to perceive the action as optimal, the goal as worthy of pursuit, etc. That is, one has instrumental reasons and these are founded on a web of non-instrumental reasons. To argue that an action is chosen because it is perceived as optimal in bringing some goal about, just produces two additional questions regarding why that particular goal is held and why that action is perceived as optimal. From this per-

spective rationality is normative in nature, in the sense that the agent pursues the goal, which according to her beliefs is good, just as she pursues it with the action, which according to her beliefs is optimal. To be rational means to act in accordance with one's well-founded beliefs which simply means that one has no strong reason not to hold a particular belief as true. To be rational, then, means to endorse the strongest reason based on the strongest evidence. But even the belief based on strong evidence may be false, implying that rationality is no guarantee of being right. The agent may draw the wrong conclusion and still rationally endorse this, simply because she holds no strong reasons not to do so, just as she may draw the right conclusion based on the wrong reasons. Right/wrong or false/true cannot be applied to distinguish the rational from the irrational, just as the incomprehensible action cannot be equated with an irrational action. The irrational action is not simply wrong – it is caused in the wrong way and, such as, an irrational action implies that something has gone wrong in the causal process or one is acting against one's better judgment. In sum, to find some principle fair/unfair is not in itself irrational, just because it deviates from, for example, principles on economic rationality – then it is simply based on a contrary set of principles and, as such, it can primary be viewed as being in opposition to economic rationality; and in a similar manner, it is not irrational to, for example, misinterpret probabilities – at most this might simply be viewed as being wrong.

Rationality is, then, not opposite to morality and neither is morality equal to irrationality. Rather, rationality implies that the individual chooses her actions in accordance with instrumental as well as social and moral principles. The individual who endorses a particular outcome distribution may then do so, because she have no strong reason not to do it and not, necessarily at least, because she finds it to be in accordance with her own principles, insofar as this might demand too much of the agent's cognitive faculties, just as it might also entail that she should be in possession of information she might not have access to. Based on this, the individual may rationally accept the consequences of a compensation system. However, implicit in this notion is also that the agent cannot produce *any* kind of strong reasons against it and, as such, this might not be reducible to factual information. In addition, rational acceptance cannot explain why social agents accepts differences at least not without taking some kind of moral or social normative foundation into consideration, just as it might not be able to explain why

these differences does not bring about a Hobbesian state of war. Rational assessment may, of course, provide some of the explanation, insofar as *knowing why* – the reason – may as argued by Folger and Martin (1986), limits the individual's ability to produce counterfactual reasons, just as it may be argued as one of respondents in my empirical study did – you do not just blame someone. Nevertheless, knowing why in itself is not enough, the individual must also accept the reasons given. Here social rationality may provide some insight, insofar as the pivotal point for this principle is to produce actions which are balanced against one's private and perhaps egoistic interests and the interests of others. However, whereas rationality and morality provide some insight into the interactional dimension of social rationality, it cannot explain why differences are accepted. To understand why differences might be accepted as fair, one has to turn to the cognitive foundation of the fair claim.

Equity theory may provide some insight into why differences might be accepted as fair, insofar as it is centered on the Aristotelian principle on fairness as relational proportionality – the fair difference is that which maintains the relational status quo. According to the early theorists on equity theory, a difference is horizontally defensible if it is based on differences in input. The reason why this difference is acceptable is simply that the social agent holds the norm on equity which governs not just the constitution of her own vertical claim, but also her horizontal claim and, as such, in order to reduce cognitive dissonance, the agent will only claim what she perceives as fair in accordance with the relational proportionality of her input. In that sense, equity theory does not solely concern what the agent wants for herself, but also what she wants for others – what Fehr and Schmidt (1999) calls social preferences. Reciprocity, as discussed above, is one type of social preference, another example might be inequity aversion. Inequity aversion states that the individual will not accept an offer which brings about an unfavorable relational difference and might even, in distribution games, decline an offer above minimum; whereas she might not, necessarily, decline an offer which brings about a favorable relational difference even though she might feel guilty accepting this. Aristotle argued that what drives behavior towards fairness is the two, sometimes opposing, principles “fairness for me” and “fairness for the other” which in the Fehr-Schmidt model is transformed into a difference between envy and guilt – whilst envy pulls towards “fairness for me”, guilt pulls towards “fairness to the other” in a similar

manner as egoism is governed by morality or sympathy. It is this ability to self-governance and to balance one's egoistic claims and envious emotions with morality, guilt and shame which characterizes social rationality. The ability to curb one's claims and to balance them with those accepted by other social agents, but also the ability to balance these claims with obligations – the claims which the agent accepts that other social agents has on her. Social rationality, in that sense, is not the opposite of egoism or equal to altruism – rather, it represents the balance between egoism and altruism – between self-regarding fairness and other-regarding fairness. It is the ability to, within reason, accept the claims of other individuals, just as it is the ability to pass claims on others within acceptable reason – within reason simply meaning that they are justifiable and that the agent herself cannot produce any strong reasons against them. From this perspective, then, the fair claim is which maintains relational proportionality; insofar as it is constituted on the relational normative foundation stating that fairness is relational proportionality.

Equity theory, then, introduces the relational dimension of fairness. But fairness is also temporally structured in the difference between past and present. The difference between past and present manifests itself as either a loss or a gain. In accordance with Aristotelian ethics and Kahneman, Knetsch, and Thaler's perspective on fairness, the cardinal rule of fairness is that no individual is allowed to gain at the expense of the other's loss. The fair difference, here, is that which maintains a temporal status quo between past and present. In a similar manner as relational proportionality binds future relational differences, past binds future. For example, from this perspective it is unfair if the employer decreases an employee's salary to obtain a higher income without a valid reason. The valid reason being the fair claim and this is the reference point – here either the employer's income or the employee's past salary. Past, in other words, binds the present through this reference point – the temporal point of comparison. Equity theory was based on social comparison. Psychophysically, the loss manifests itself as more painful than a possible gain and, as such, one should expect negative deviations from the past to be associated with significant negative emotional experience, while positive deviations to be associated with less significant positive emotional experience. This assumption on changes in hedonic experience follows the assumption on hedonic adaptation stating individuals over time adapt their preferences to the situation facing them.

This point of hedonic experience to which the individual has adapted, then, becomes what is perceived as normal, as the point of comparison. From this perspective, if the unfair is perceived as a loss, the unfair is also associated with painful hedonic experience. That is, however, not to argue that the painful is always unfair or that the pleasurable or neutral experience is always fair. From a first-person perspective, hedonic experience is problematic to apply as the foundation of fairness, insofar as the agent, on the one hand, may adapt her preferences to unfair conditions, just as she, on the other hand, might also counter-adapt her preferences to fair conditions – the former principle may be illustrated by the fable about the fox and the sour grapes, whereas the latter principle may be illustrated by the proverb “the grass is always greener...”.

In my empirical study, I found that a relational difference which was perceived as vertically and horizontally fair, from a third-person perspective, is associated with a positive emotional experience in the receiver, while it is associated with a neutral emotional experience in the non-receiver regardless of past relational status quo. In that sense, from a third-person perspective and in cases when the difference is perceived as vertically and horizontally fair, status quo, be that either relational or temporal, may not matter as much as earned claims. As one of the respondents argued, the receiving-agent did something extra. However, as also argued in the interviews, even the agent who obtains the reward in a fair manner is not allowed to hurt the other – in the sense that the agent should not brag about his/her bonus, just as the non-receiving agent should not blame the receiving agent, insofar as the other cannot be held accountable for bringing about the after-situation, at least not from a causal perspective. From this it seems to follow that the principles of social rationality may not necessarily encompass the maintenance of a relational and temporal status quo, at least, not from a third-person perspective and in cases in which the after-situation is perceived as vertically and horizontally fair. Social rationality, on the other hand, encompasses how the agents should act towards one another. Hence, the principles on social rationality do not prevent the social agents from pursuing their private goals – it solely states that this should be done in a manner which does harm the other – occurs at the expense of the other.

As argued in the previous section, the non-receiving agent should not blame the receiving-agent for matters which are not within her control. This perspective emerges

from the theoretical connection between causality of blame and the causality of the rewarding processes. In accordance with Folger and Cropanzano's (1998; 2001) theory on fairness as blame, from a vertical perspective, the manager or the organization can be held directly responsible for bringing about a particular outcome-situation. The judgment on fairness is here based on a comparison between the actual world and the counterfactual or possible world, which should have emerged had, for example, the manager acted in accordance with the governing relational normative expectations. The unfair outcome-situation is the situation which deviates from that which is perceived as being in accordance with cognitive normative expectations. Thus, the fair claim is also constituted on a conception of normality and the judgment based on a comparison between the actual world and the counterfactual world which is in accordance with normative expectations. From a horizontal perspective, however, the receiving agent or the agent who differentiates herself from her colleagues cannot be blamed, at least not from a causal perspective for bringing about some particular outcome-state, insofar as she cannot decide whether her actions are rewarded. The causal link is then indirect, because it necessitates the mediation of management. This might also explain why the employees who differentiate themselves from their colleagues are not blamed for bringing about some particular outcome state, but solely for acting in an antisocial manner or producing actions which are not in accordance with the governing relational norms. While the non-receiving agent might not be able to blame the receiving agent for bringing about some particular outcome-state, she might be able to blame herself for not acting in a particular manner and here a direct causal relation may be established. In one of cases presented to the respondents in my empirical study, the non-receiving agent had the opportunity of bringing some particular state about, but declined it and as one of the interviewees argued, she has only herself to blame. Causal agency may, in other words, matter, at least direct causal agency.

In accordance with what has been argued so far, the fair claim is constituted by normative expectations related with the governing relational norms – such as reciprocity – just as it is constituted by relational, temporal, and cognitive normative expectations related with social comparisons, comparisons between before and after, and comparisons between the actual and the counterfactual world. Besides these matters which constitute the fair outcome claim, it has also been argued that horizontal relations are based

on the principle of social rationality – related with the moral and social limits of behavior and desire. Hence, from this perspective, rational acceptance may not be the only reason to why some differences are accepted, insofar as what seems to matter is whether these are in accordance with the complex set of norms governing not only the outcome, but also the causal process which brings this outcome about, in the sense that while it may be permissible to be ambitious and egoistic, one is not allowed to inflict harm on others, just as some actions might not be in accordance with the governing moral and social norms. From this perspective, a more complex notion of rationality emerges, a notion of rationality which cannot be reduced to instrumental rationality. The argument here is not, however, that compensation systems as the one introduced in a previous section cannot work. Rather, the argument is simply that it depends on the vertical and horizontal normative foundation.

From this discussion of the nature and the constitution of the fair claim, I now turn to the third question on whether acts of differentiation might not undermine the motivational base of the compensation system if this destroys social bonds between the agents which might lead to a loss of belongingness which, subsequently, might lead to a loss of motivation to achieve. From what has been argued above, it seems to follow that relational proportionality, temporality, and cognitive normative expectations all produces bounds which binds future expectations, just as it have been argued that due to the constitution of the social world and the social and moral expectations which emerges solely from taking part in this social world, social relations also binds interaction and action, because it forces the social agent to balance her right to claim with her obligations, just as it forces her to balance her private interests with the interests of other individuals. The force which these considerations impel themselves on the agent may either be due to matters related with esteem/self-esteem, costs and benefits, sympathy, or a self-commitment to a felt obligation. However, from a, perhaps, more basic and less negative perspective, it might simply be argued that, at least, in some cases, social agents adheres to social norms, respect moral norms and social ties due to a need to belong. Durkheim (1915/2008) argued that morality emerged from this need to belong – a need to feel that one is part of something transcending one's present and material life. In a similar manner as the Smith-Boudon model centered on the notion of sympathy or the Kantian model based on the notion of a kingdom of reason from which the moral obli-

gation emerges, insofar as being governed by one's own will does not imply a liberation from one's other-regarding responsibility, rather it emerges from this – from being the producer of an action, from willing to bring some state about also follows a will to the action which might bring this state about, and a will to take on oneself the responsibility which flows from willing the goal and the action. Regardless of philosophical perspective, then, it seems that social life entails, by necessity, some kind of either moral responsibility or consideration emerging from either being interconnected through consequences, will, socioemotional structuring, or simply through a social need. According Maslow (1965/2011), the need-deficient agent primarily desires the company of others in order to satisfy her social need, she perceives these other agents solely as means to her own end. This kind of social relation is asymmetrical in nature because it is based one of the agents wanting something from the other and, as such, this negative relation centered on dependence prevents the agent being an autonomous agent who engage with the world in an autonomous manner and solely commits herself to others by will – only then can the Kantian imperative on respect/dignity be revealed in its true sense, as the imperative which commands the autonomous agent to respect herself as well as others by not treating these others as means to her goal or letting herself be treated as a means to somebody else goal. Because only at this psychological state, is a true sense of self-esteem invoked as a true notion of non-contingent self-worth, just as only at this state is the agent able to realize her true desires, her true self. Satisfaction of this social need, then, is necessary in order to bring about this self-realizing autonomous social agent who interacts with the world in an autonomous manner and who is driven by a desire to achieve and develop herself. According to Deci and Ryan (2000), only at this psychological state is the autonomous agent dependent by choice – motivated by a true commitment to others, in the sense that she chooses to choose to be dependent, chooses to choose the other. Hence, autonomous agency and self-realization is not opposite to dependency, rather it is contingent upon the satisfaction of this basic psychological need to belong as also argued by Baumeister and Leary (1995) in their classic article on the psychological consequences of social exclusion. Recent research has also demonstrated that this need to belong and to be accepted by others is so basic that the respondents in the experiments might even react negatively to being excluded by a non-human agent, just as it has also demonstrated that satisfaction of this psychological need is necessary

to morality, insofar as subjects who suffers social exclusion might feel less human and perceive others in a similar manner – thus, exclusion might be experienced as exclusion from the humankind, as being stripped of one’s humanity or dignity, or ostracized from the Kantian kingdom of reason. From this state of psychological anxiety follows a kind of moral anomie similar in kind to Merton’s (1938/1968) notion of social anomie which follows from relative deprivation – a state of cognitive dissonance when one is prevented from obtaining what one feels entitled to or at least as entitled those who obtains it. However, before reaching this dystopian state, social exclusion or fear of this might lead the social agent to be highly sensitive to social norms and, as some research has demonstrated, the higher the need to belong, the higher the sensitivity to social norms and to the opinions of others, hence the more fierce the grip of the norm on the mind of the agent is.

From this it seems to follow that a compensation system seeking to overcome possible social norms preventing acts on differentiation by rewarding this behavior, might have the adverse effect since the agents violating the norms may become more sensitive to them, just as it might undermine the social relations which might lead to need-deficiency and, as such, the system might undermine the motivational foundation of individuality and differentiation. In my small empirical study, I found that a vertically and horizontally fair after-situation is more likely to produce a neutral emotional experience in the non-receiving agent, just as it is more likely to produce a neutral relational influence than a negative and positive one. The relational influence variable is positively correlated with the emotional experience of the non-receiving; hence one should expect that in situations producing highly negative emotional experience in the non-receiving, the relational influence of the after-situation would also be negative. Although correlated, the two variables are far from perfectly correlated which might be associated with, on the one hand, the problem of indirect causality discussed above, and, on the other hand, it might also be related with the principle of social rationality preventing the non-receiving social agent from blaming the receiving agent and the receiving agent from, for example, bragging about her/his favorable outcome. That is, the less than perfect correlation or lagging effect of emotions on relational influence might be explained by either social rationality or a lack of direct logical causal relationship. To some extent this result differs from the Fehr-Schmidt model which would suggest a

higher correlation due to reciprocity, however, the lack of logical causal agency might be generating the difference, insofar as in distribution games a direct causal relationship can be established between the actions of the distributor and the after-situation she tries to bring about. This logical relationship is not possible to establish in cases related with compensation due to the difference between the causality of rewarding and the causality of blame. Hence, due to this difference in causal logic and the fact that only an indirect causal relation can be established, norms might primarily target actions as being antisocial not for being possible means to bringing about some particular after-situation.

In conclusion, then, it can be argued that fairness is more than rational acceptance, insofar as fairness is related with the normative foundation on which vertical and horizontal relations are centered. Vertically, the fair claim is not reducible to matters related with outcome, but encompasses a behavioral – non-consequentialistic dimension. Horizontally, the fair claim is primarily non-consequentialistic, because it is related with matters on what constitutes “fairness for me” and “fairness for the other”. The asocial nature of fair claim constituted in contemporary compensation systems may not capture this underlying social and moral foundation of the relational structuring of the organization and as a consequence, the compensation system might undermine the very effect it is designed to bring about, insofar as it might change the underlying normative foundation of vertical relations (the non-consequentialistic dimension), just as it might be unaligned with the three normative constitutive dimensions of the fair claim, and as a consequence it might be judged as bringing about unfair consequences. Furthermore, because the theories does not recognize the relational nature of social life or its interconnectedness, the compensation system might destroy or damage these relations by motivating behavior which is unaligned with the horizontal social and moral normative foundation and in cases when the after-situation is perceived as vertically unfair and as a product of horizontal norm violating behavior, the relational influence might be negative and as a consequence of this, the horizontal social relations might be damaged leading to a lower degree of satisfaction of the need to belong and this need-deficiency might undermine the motivation to achieve, to excel, to differentiate oneself.

## References

- Adams, J.S. (1963): Toward an Understanding of Inequity. *Journal of Abnormal and Social Psychology*, 67/5, pp. 422-436
- Akerlof, G. (1982): Labor Contracts as Partial Gift Exchange. *Quarterly Journal of Economics*, 97/4, pp. 543-569
- Allison, H. (1990): *Kant's Theory of Freedom*. Cambridge University Press
- Ambrose, M. and Schminke, M. (2009): The Role of Overall Justice Judgments in Organizational Justice Research: A Test of Mediation. *Journal of Applied Psychology*, 94/2, pp. 491-500
- Andersen, L.B. (2012): Styring af den offentlige sektor: Incitament, motivation og normer. *Produktivitets Kommissionen, Baggrundspapir – Trivsel og produktivitet*
- Andersen, L.B. and Pallesen, T. (2008): "Not Just for the Money?" How Financial Incentives Affect the Number of Publications at Danish Research Institutions. *International Public Management Journal*, 11/1, pp. 28-47
- Ariely, D. (2009): *Predictable Irrational (Revised and Expanded Edition)*. Harper
- Ariely, D. and Heyman, J. (2004): Effort for Payment: A Tale of Two Markets. *Psychological Science*, 15/11, pp. 787-793
- Aristoteles (350BC/2007a): *Retorik*. Museum Tusulanums forlag
- Aristoteles [Aristotle] (350BC/2007b): *Etikken [Nicomachean Ethics]*. Det lille forlag
- Armstrong, M.; Cummings, A.; Hastings, S.; and Wood, W. (2003): *Job Evaluation: A Guide to Achieving Equal Pay*. Kogan Page LTD
- Atzmüller, C. and Steiner, P.M. (2010): Experimental Vignette Studies in Survey Research. *Methodology*, 6/5, pp. 128-138
- Axelrod, R. (2006): *The Evolution of Cooperation*. Basic Books
- Baier, A. (1994): *A Progress of Sentiments*. Harvard University Press
- Balaguer, M. (1998): *Platonism and Anti-Platonism in Mathematics*. Oxford University Press
- Bandiera, O.; Barankay, I.; and Rasul, I. (2005): Social Preferences and Response to Incentives: Evidence from Personnel Data. *The Quarterly Journal of Economics*, 120/3, pp. 917-962
- Barnard, C. (1938/1968): *The Functions of the Executive*. Harvard University Press
- Barney, J. (1991): Firm Resources and Sustained Competitive Advantage. *Journal of Management*, 17/1, pp. 99-120

- Baron, J.N. and Pfeffer, J. (1994): The Social Psychology of Organizations and Inequality. *Journal of Psychology Quarterly*, 57/3, pp. 190-209
- Bartels, J.M.; Baumeister, R.; Ciarocco, N.J.; DeWall, C.N.; and Twenge, J.M. (2007): Social Exclusion Decreases Prosocial Behavior. *Journal of Personality and Social Psychology*, 92/1, pp. 56-66
- Barth, E.; Bratsberg, B.; Hægeland, T.; and Raaum, O. (2012): Performance Pay, Union Bargaining and Within-firm Wage Inequality. *Oxford Bulletin of Economics and Statistics*, 74/3, pp. 327-362
- Bastian, B. and Haslam, N. (2010): Excluded from Humanity: The Dehumanizing Effects of Social Ostracism. *Journal of Experimental Social Psychology*, 46, pp. 107-113
- Bastian, B.; Jetten, J.; Chen, H.; Radke, H.R.M; Harding, J.F.; and Fasoli, F. (2013): Losing our Humanity: The Self-Dehumanizing Consequences of Social Ostracism. *Personality and Social Psychology Bulletin*, 39/2, pp. 156-169
- Baumeister, R.F. and Leary, M.R. (1995): The Need to Belong: Desire for Interpersonal Attachments as a Fundamental Human Motivation. *Psychological Bulletin*, 117/3, pp-497-529
- Bazerman, M.H.; Loewenstein, G.F.; and Thompson, L. (1989): Social Utility and Decision Making in Interpersonal Contexts. *Journal of Personality and Social Psychology*, 57/3, pp. 426-441
- Bazerman, M.H.; Loewenstein, G.F.; and White, S.B. (1992): Reversals of Preference in Allocation Decisions: Judging an Alternative versus Choosing among Alternatives. *Administrative Science Quarterly*, 37, pp. 220-240
- Bazerman, M.H; White, S.B.; Loewenstein, G.F. (1995): Perceptions of Fairness in Interpersonal and Individual Choice Situations. *Current Directions in Psychological Science*, 4/2, pp. 39-43
- Beck, U. (2007): *Schöne Neue Arbeitswelt*. Suhrkamp
- Becker, B. and Huselid, M. (2006): Strategic Human Resource Management: Where Do We Go From Here? *Journal of Management*, 32/6, pp. 898-925
- Beller, S. (2008): Deontic Norms, Deontic Reasoning, and Deontic Conditionals. *Thinking and Reasoning*, 14/4, pp. 305-341
- Beller, S. (2010): Deontic Reasoning Reviewed: Psychological Questions, Empirical Findings, and Current Theories. *Cognitive Processing*, 11, pp. 123-132
- Bentham, J. (1780/2007): *An Introduction to the Principles of Morals and Legislation*. Dover Philosophical Classics
- Berezin, M. (2005): Emotions and the Economy. In. Smelser, N.J. and Swedberg, R. (ed.) (2005): *The Handbook of Economic Sociology*. Russell Sage Foundation, pp. 109-130

- Berger, L.A. and Berger, D.R. (2008): *The Compensation Handbook*. McGraw Hill
- Berger, P. and Luckmann, T. (1966/2002): *Den Samfundsskabte Virkelighed* [Eng. The Social Construction of Reality]. Lindhardt og Ringhof
- Bewley, T. (1998): Why Not Cut Pay?. *European Economic Review*, 42, 459-490
- Bewley, T. (2004): Fairness, Reciprocity, and Wage Rigidity. *IZA Discussion Paper*
- Bhaskar, R. (1978/2008): *A Realist Theory of Science*. Verso
- Bhaskar, R. (1986/2009): *Scientific Realism and Human Emancipation*. Routledge
- Bhaskar, R. (1989/1998): *The Possibility of Naturalism*. Routledge
- Bicchieri, C. (2006): *The Nature and Dynamics of Social Norms*. Cambridge University Press
- Bies, R. and Shapiro, D. (1987): Interactional Fairness Judgments: The Influence of Causal Accounts. *Social Justice Research*, 1/2, pp. 199-218
- Binmore, K. (2009): *Rational Decisions*. Princeton University Press
- Boudon, R. (1976): Comment on Hauser's Review of Education, Opportunity, and Social Inequality. *American Journal of Sociology*, 81/5, pp. 1175-1187
- Boudon, R. (1996): The Cognitivist Model: A Generalized Rational-Choice Model. *Rationality and Society*, 8/2, pp. 123-150
- Boudon, R. (2001): *The Origin of Values: Sociology and Philosophy of Beliefs*. Transaction Publishers
- Boudon, R. (2006): Homo-Sociologicus: Neither a Rational nor Irrational Idiot. *Papers: Revista De Sociologia*, 80, pp. 149-169
- Boudon, R. (2011): Ordinary Rationality: The Core of Analytical Sociology. In: Demeulenaere, P. (ed.) (2011): *Analytical Sociology and Social Mechanisms*. Cambridge University Press, pp. 33-49
- Bradburn, N.; Sudman, S.; and Wansink, B. (2004): *Asking Questions*. Jossey-Bass
- Burns, J. M. (1978): *Leadership*. Perennial
- Camerer, C.F. (2003): *Behavioral Game Theory*. Russell Sage Foundation
- Carrell, M.R. and Dittrich, J.E. (1978): Equity Theory: The Recent Literature, Methodological Considerations, and New Directions. *The Academy of Management Review*, 3/2, pp. 202-210
- Cinyabuguma, M.; Page, T.; and Putterman, L. (2006): Can Second-Order Punishment Deter Perverse Punishment?. *Experimental Economics*, 9, pp. 265-279
- Clark, M. S. (1984): Record Keeping in Two Types of Relationships. *Journal of Personality and Social Psychology*, 47/3, pp. 549-557

- Clark, V.L.; Creswell, J.; Green, D.; and Shope, R. (2010): Mixing Quantitative and Qualitative Approaches. In: Hesse-Biber, S. and Leavy, P. (ed.) (2010): *Handbook of Emergent Methods*. The Guilford Press, pp. 363-387
- Coase, R. H. (1937): The Nature of the Firm. *Economica*, 4/16, pp. 386-406
- Coleman, J.S. (1990): *Foundations of Social Theory*. The Belknap Press of Harvard University Press
- Coleman, J.S. (1993a): The Role of Rights in a Theory on Social Action. *Journal of Institutional Economics*, 149/1, pp. 213-232
- Coleman, J.S. (1993b): The Design of Organizations and the Right to Act. *Sociological Forum*, 8/4, pp. 527-546
- Colquitt, J.; Noe, R.; and Jackson, C. (2002): Justice in Teams: Antecedents and Consequences of Procedural Justice Climate. *Personnel Psychology*, 55/1, pp. 83-105
- Colquitt, J.A.; Greenberg, J.; and Zapata-Phelan, C.P. (2005): What is Organizational Justice? A Historical Overview. In: Greenberg, J. and Colquitt, J.A. (2005): *Handbook of Organizational Justice*, Lawrence Erlbaum Associates, pp. 3-56
- Connelly, B.L.; Tihanyi, L.; Crook, R.; and Gangloff, K.A. (2013): Tournament Theory: Thirty Years of Contests and Competitions. *Journal of Management*, 20/10, pp. 1-32
- Cook, K.S. and Hegtvedt, K.A. (1983): Distributive Justice, Equity, and Equality. *Annual Review of Sociology*, 9, pp. 217-241
- Corby, S; Palmer, S.; Linop, E. (2008): Trends and Tensions: an overview. In: Corby, S; Palmer, S.; Linop, E. (ed.) (2008): *Rethinking Reward*. Palgrave Macmillian
- Cremer, D.D. and Tenbrunsel, A.E. (ed.) (2012): *Behavioral Business Ethics*. Routledge
- Creswell, J. (2003): *Research Design*. Sage Publications
- Creswell, J. (2007): *Qualitative Inquiry and Research Design*. Sage Publications
- Creswell, J.W.; Clark, V.L.P; Gutmann, M.L.; and Hanson, W.E. (2002): Advanced Mixed Methods Research Designs. In: Tashakkori, A. and Teddlie, C. (ed.) (2003): *Handbook of Mixed Methods in Social and Behavioral Research*. Sage Publications, pp. 241-272
- Cropanzano, R. and Ambrose, M.L. (2001): Procedural and Distributive Justice Are more Similar than You Think: A Monistic Perspective and a Research Agenda. In: Greenberg, J. and Cropanzano, R. (ed.) (2001): *Advances in Procedural Justice*. Stanford University Press, pp. 119-151
- Cropanzano, R.; Byrne, Z.; Bobocel, D.R.; and Rupp, D. (2001): Moral Virtues, Fairness Heuristics, Social Entities, and Other Denizens of Organizational Justice. *Journal of Vocational Behavior*, 58, pp. 164-209

- Cropanzano, R.; Goldman, B.; and Folger, R. (2003): Deontic Justice: The Role of Moral Principles in Workplace Fairness. *Journal of Organizational Behavior*, 24/8, pp. 1019-1024
- Danermark, B.; Ekström, M.; Jakobsen, L.; and Karlsson, J.Ch. (2002): *Explaining Society: Critical Realism in the Social Sciences*. Routledge
- Darrington, J.W. and Howell, G.A. (2011): Motivation and Incentives in Relational Contracts. *Journal of Financial Management of Property and Construction*, 16/1, pp. 42-51
- Davidson, D. (1963/2001): Actions, Reasons, and Causes. In: Davidson, D. (2001): *Essays on Actions and Events*. Clarendon Press
- Davidson, D. (1978/2010): Paradoxes of Irrationality. In: Davidson, D. (2010): *Problems of Rationality*. Clarendon Press
- Davidson, D. (2001): *Essays on Actions and Events*. Clarendon Press
- Davidson, D. (2010): *Problems of Rationality*. Clarendon Press
- Deci, E. and Ryan, R. (1995): Human Autonomy – The Basis for True Self-Esteem. In: Kernis, M. (ed.) (1995): *Efficacy, Agency, and Self-Esteem*. Springer Science
- Deci, E.L. and Ryan, R.M. (1985): *Intrinsic Motivation and Self-Determination in Human Behavior*. Kluwer Academics
- Deci, E.L. and Ryan, R.M. (2000): The “Whats” and “Why” of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry*, 11/4, pp. 227-268
- Demeulenaere, P. (ed.) (2011): *Analytical Sociology and Social Mechanisms*. Cambridge University Press
- Deming, W.E. (1982/2000): *Out of the Crisis*. The MIT Press
- Descartes, R. (1637/1998): *Discourse on Method*. Hackett Publishing Company
- Durkheim, E. (1897/1951): *Suicide – A Study in Sociology*. The Free Press
- Durkheim, E. (1915/2008): *The Elementary Forms of the Religious Life*. Dover Publications
- Elster, J. (1977): Ulysses and the Sirens: A Theory of Imperfect Rationality. *Social Science Information*, 16/5, pp. 469-526
- Elster, J. (1983): *Sour Grapes*. Cambridge University Press
- Elster, J. (1989a): *Solomonic Judgments*. Cambridge University Press
- Elster, J. (1989b): *The Cement of Society: A Study of Social Order*. Cambridge University Press

- Elster, J. (1992): *Local Justice: How Institutions Allocate Scarce Goods and Necessary Burdens*. Russell Sage Foundation
- Elster, J. (1998): Emotions and Economic Theory. *Journal of Economic Literature*, 35/1, pp. 47-74
- Elster, J. (1999): *Alchemies of the Mind: Rationality and the Emotions*. Cambridge University Press
- Elster, J. (2005): Fehr on Altruism, Emotion, and Norms. *Analyse & Kritik*, 27, pp. 197-211
- Elster, J. (2006): Fairness and Norms. *Social Research*, 73/2, pp. 365-376
- Elster, J. (2007): *Explaining Social Behavior: More Nuts and Bolts for the Social Sciences*. Cambridge University Press
- Elster, J. (2009a): Emotions. In: Hedström, P. and Bearman, P. (ed.) (2009): *The Oxford Handbook of Analytical Sociology*. Oxford University Press, pp. 51-71
- Elster, J. (2009b): Norms. In: Hedström, P. and Bearman, P. (ed.) (2009): *The Oxford Handbook of Analytical Sociology*. Oxford University Press, pp. 195-217
- Elster, J. (2009c): *Reason and Rationality*. Princeton University Press
- Elster, J. (2010): Self-poisoning of the Mind. *Philosophical Transactions of the Royal Society B*, 365, pp. 221-226
- Elster, J. (2011): Indeterminacy of Emotional Mechanisms. In: Demeulenaere, P. (ed.) (2011): *Analytical Sociology and Social Mechanisms*. Cambridge University Press, pp. 50-63
- Falk, A.; Fehr, E.; and Fischbacher, U. (2003): On the Nature of Fair Behavior. *Economic Inquiry*, 41/1, pp. 20-26
- Falk, A.; Fehr, E.; and Fischbacher, U. (2005): Driving Forces behind Informal Sanctions. *Econometrica*, 73/6, pp. 2017-2030
- Faulkender, M.; Kadyrzhanova, D.; Prabhala, N.; and Senbet, L. (2010): Executive Compensation: An Overview of Research on Corporate Practices and Proposed Reforms. *Journal of Applied Corporate Finance*, 22/1, pp. 107-118
- Fehr, E. and Falk, A. (2002): Psychological Foundations of Incentives. *European Economic Review*, 45, pp. 687-724
- Fehr, E. and Fischbacher, U. (2002): Why Social Preferences Matter – The Impact of Non-Selfish Motives on Competition, Cooperation, and Incentives. *The Economic Journal*, 112, C1-C33
- Fehr, E. and Gächter, S. (2000): Fairness and Retaliation: The Economics of Reciprocity. *Journal of Economic Perspectives*, 14/3, pp. 159-181

- Fehr, E. and Gächter, S. (2002): Do Incentive Contracts Undermine Voluntary Cooperation?. *University of Zurich*, Working Paper No. 34
- Fehr, E. and Schmidt, K.M. (1999): A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics*, pp. 817-858
- Fehr, E.; Fischbacher, U.; and Gächter, S. (2001): Strong Reciprocity, Human Cooperation, and the Enforcement of Social Norms. *Human Nature*, 13/1, pp. 1-25
- Fehr, E.; Gächter, S.; and Kirschsteiger, G. (1997): Reciprocity as a contract enforcement device: Experimental Evidence. *Econometrica*, 65/4, pp. 833-860
- Fehr, E.; Klein, A.; and Schmidt, K.M. (2007): Fairness and Contract Design. *Econometrica*, 75/1, pp. 121-154
- Festinger, L. (1954): A theory on Social Comparison Processes. *Human Relations*, 7, pp. 117-140
- Festinger, L. (1957/1985): *A Theory of Cognitive Dissonance*. Stanford University Press
- Finch, J. (1987): Research Note: The Vignette Technique in Survey Research. *Sociology*, 21/1, pp. 105-114
- Fish, J.S. (2013): Homo-duplex revisited: A defense of Emile Durkheim's theory of the moral self. *Journal of Classical Sociology*, 13/3, pp. 338-358
- Fiske, A.P. (1991): *Structures of Social Life: The Four Elementary Forms of Human Relations*. Free Press
- Fleetwood, S. (2005): Ontology in Organization and Management Studies: A Critical Realist Perspective. *Organization*, 12/2, pp. 197-222
- Folger, R. (1998): Fairness as Moral Virtue. In: Schminke, M. (ed.) (1998): *Managerial Ethics: Moral Management of People and Processes*. Lawrence Erlbaum, Inc., pp. 13-34
- Folger, R. (2012): Deonance: Behavioral Ethics and Moral Obligation. In: Cremer, D.D. and Tenbrunsel, A.E. (ed.) (2012): *Behavioral Business Ethics*. Routledge, pp. 123-142
- Folger, R. and Cropanzano, R. (1998): *Organizational Justice and Human Resource Management*. Sage Publications
- Folger, R. and Cropanzano, R. (2001): Fairness Theory: Justice as Accountability. In: Greenberg, J. and Cropanzano, R. (ed.) (2001): *Advances in Organizational Justice*. Stanford University Press, pp. 1-55
- Folger, R. and Martin, C. (1986): Relative Deprivation and Referent Cognitions: Distributive and Procedural Justice Effects. *Journal of Experimental Social Psychology*, 22, pp. 531-546
- Folger, R. and Skarlicki, D. (2008): The Evolutionary Bases of Deontic Justice. In: Gilliland, S.W.; Steiner, D.D.; and Skarlicki, D.P. (ed.) (2008): *Justice, Morality, and Social Responsibility*. Information Age Publishing Inc., pp. 29-62

- Folger, R.; Ganegoda, D.; Rice, D.; Taylor, R.; Wo, D.X. (2013): Bounded Autonomy and Behavioral Ethics: Deonance and Reactance as Competing Motives. *Human Relations*, 66/7, pp. 905-924
- Frank, R. (1985): *Choosing the Right Pond – Human Behavior and the Quest for Status*. Oxford University Press
- Frey, B. (1997): *Not Just for the Money*. Edward Elgar Publishing
- Fudge, R.S. and Schlacter, J.L. (1999): Motivating Employees to Act Ethically: An Expectancy Theory Approach. *Journal of Business Ethics*, 18/3, pp. 295-304
- Gächter, S. and Riedl, A. (2005): Moral Property Rights in Bargaining with Infeasible Claims. *Management Science*, 51/2, pp. 249-263
- Gächter, S.; Kessler, E.; and Königstein, M. (2008): Performance Incentives and the Dynamics of Voluntary Cooperation. *Symposium zur Ökonomischen Analyse der Unternehmung*
- Gächter, S.; Nosenzo, D.; and Sefton, M. (2012): The Impact of Social Comparison on Reciprocity. *Scandinavian Journal of Economics*, 114/4, pp. 1346-1369
- Gaus, G. F. (2008): *On Philosophy, Politics, and Economics*. Wadsworth Cengage Learning
- Gergen, K.J.; Greenberg, M.; and Willis, R.H. (ed.) (1980): *Social Exchange*. Plenum Press
- Gill, D. and Stone, R. (2010): Fairness and Desert in Tournaments. *Games and Economic Behavior*, 69, pp. 346-364
- Gilliland, S.W.; Steiner, D.D.; and Skarlicki, D.P. (ed.) (2008): *Justice, Morality, and Social Responsibility*. Information Age Publishing Inc.
- Gilovich, T.; Griffin, D.W.; and Kahneman, D. (ed.) (2002): *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press
- Giorgi, A. (2009): *The Descriptive Phenomenological Method in Psychology*. Duquesne University Press
- Gneezy, U. and Rustichini, A. (2000): A Fine is a Price. *The Journal of Legal Studies*, 29/1, pp. 1-17
- Greenberg, J. (1990): Organizational Justice: Yesterday, Today, and Tomorrow. *Journal of Management*, 16/2, pp. 399-432
- Greenberg, J. and Colquitt, J.A. (2005): *Handbook of Organizational Justice*, Lawrence Erlbaum Associates
- Greenberg, J. and Cropanzano, R. (ed.) (2001): *Advances in Organizational Justice*. Stanford University Press

- Greenberg, J. and Cropanzano, R. (ed.) (2001): *Advances in Procedural Justice*. Stanford University Press
- Greene, J. and Hall, J. (2010): Dialectics and Pragmatism. In: Tashakkori, A. and Teddlie, C. (ed.) (2010): *Sage Handbook of Mixed Methods in Social and Behavioral Research*. Sage Publications, pp. 119-144
- Greenspan, P. (2007): Practical Reasons and Moral “Ought”. In: Shafer-Landau, R. (ed.) (2007): *Oxford Studies in Metaethics: Vol. 2*. Oxford University Press, pp. 172-194
- Gruman, J. and Saks, A. (2011): Performance Management and Employee Engagement. *Human Resource Management Review*, 21/2, pp. 123-136
- Haidt, J. (2012): *The Righteous Mind*. Pantheon Books
- Haidt, J. Seder, P., and Kesebir, S. (2008): Hive Psychology, Happiness, and Public Policy. *Journal of Legal Studies*, 37, pp. S133-S156
- Hamilton, E. and Cairns, H. (ed.) (1961): *Plato – The Collected Dialogues*. Princeton University
- Hammersley, M (2013): *The Relationship Between Qualitative and Quantitative Research: Paradigm Loyalty Versus Methodological Eclecticism*.  
<http://martyhammersley.files.wordpress.com/2013/03/relationship-between-qualitative-and-quantitative-research.pdf> [Accessed 12.12.2013]
- Hechter, M. and Opp, K-D. (ed.) (2001): *Social Norms*. Russell Sage Foundation
- Hedström, P. (2005): *Dissecting the Social*. Cambridge University Press
- Hedström, P. and Bearman, P. (ed.) (2009): *The Oxford Handbook of Analytical Sociology*. Oxford University Press
- Hedström, P. and Udehn, L. (2011): Analytical Sociology and Theories of the Middle-Range. Pp. 25-50
- Heider, F. (1946): Attitudes and Cognitive Organization. *The Journal of Psychology*, 21, pp. 107-112.
- Herzberg, F. (1968/1987): One More Time: How Do You Motivate Employees?. *Harvard Business Review*, September-October, pp.5-16
- Herzberg, F.; Mausner, B.; and Snyderman, B.B. (1959/2010): *The Motivation to Work*. Transaction Publishers
- Hesse-Biber, S. and Leavy, P. (ed.) (2010): *Handbook of Emergent Methods*. The Guilford Press
- Hicks, J. (1974): *The Crisis of Keynesian Economics*. Basic Books – INC Publishers
- Hobbes, T. (1651/1985): *Leviathan*. Penguin Classics

- Homans, G.C. (1954): The Cash Posters – A Study of a Group of Working Girls. *American Sociological Review*, 19/6, pp. 724-733
- Homans, G.C. (1958): Social Behavior as Exchange. *American Journal of Sociology*, 63/6, pp. 597-606
- Hondeghem, A. and Perry, J.L. (2009): EGPA Symposium on Public Service Motivation and Performance: Introduction. *International Review of Administrative Sciences*, 75/5, pp. 5-9
- Hume, D. (1739/1985): *A Treatise on Human Nature*. Penguin Classics
- Hume, D. (1748/1975): *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*. Clarendon Press
- Hume, D. (1779/1947): *Dialogues Concerning Natural Religion*. Library of Liberal Arts
- Huselid, M. and Becker, B. (2011): Bridging Micro and Macro Domains: Workforce Differentiation and Strategic Human Resource Management. *Journal of Management*, 37/2, pp. 421-428
- Huseman, R.C.; Hatfield, J.D.; and Miles, E.W. (1987): A New Perspective on Equity Theory: The Equity Sensitivity Construct. *The Academy of Management Review*, 12/2, pp. 222-234
- Ingster, B. (2008): Job Analysis, Documentation, and Job Evaluation. In: Berger, L. and Berger, D. (ed.) (2008): *The Compensation Handbook*. McGraw Hill, pp. 95-110
- Irwin, K. and Horne, C. (2013): A Normative Explanation of Antisocial Punishment. *Social Science Research*, 42, pp. 562-570
- Isaac, R.G.; Zerbe, W.L.; and Pitt, D.C. (2001): Leadership and Motivation. *Journal of Managerial Issues*, 13/2, pp. 212-226
- Jamieson, S. (2004): Likert Scales: how to (ab)use them. *Medical Education*, 38, pp. 1217-1218
- Jaques, E. (1961): *Equitable Payment*. Heinemann
- Jasso, G. (2006): Factorial Survey Methods for Studying Beliefs and Judgments. *Sociological Methods and Research*, 34/1, pp. 334-423
- Kahn, C. H. (1987): Plato's Theory of Desire. *The Review of Metaphysics*. 41/1, pp. 77-103
- Kahneman, D and Tversky, A. (1973): Availability: A Heuristic for Judging Frequency and Probability. In: Kahneman, D.; Slovic, P.; and Tversky, A. (ed.) (1982): *Judgment under Uncertainty: Heuristics and Biases*. Cambridge University Press, pp. 163-179
- Kahneman, D and Tversky, A. (1979): The Simulation Heuristic. In: Kahneman, D.; Slovic, P.; and Tversky, A. (ed.) (1982): *Judgment under Uncertainty: Heuristics and Biases*. Cambridge University Press, pp. 201-211

- Kahneman, D. (1999): Objective Happiness. In: Kahneman, D; Diener, E.; and Schwarz, N. (ed.) (1999): *Well-Being: The Foundations of Hedonic Psychology*. Russell Sage Foundation, pp. 3-25
- Kahneman, D. (2000a): Experienced Utility and Objective Happiness: A Moment-Based Approach. In: Kahneman, D. and Tversky, A. (ed.) (2000): *Choices, Values, and Frames*. Cambridge University Press, Pp. 673-692
- Kahneman, D. (2000b): Evaluation by Moments: Past and Future. In: Kahneman, D. and Tversky, A. (ed.) (2000): *Choices, Values, and Frames*. Cambridge University Press, Pp. 693-708
- Kahneman, D. (2011): *Thinking, Fast and Slow*. Penguin Psychology
- Kahneman, D. and Miller, D. T. (1986): Norm Theory: Comparing Reality to its Alternatives. In: Gilovich, T.; Griffin, D.W.; and Kahneman, D. (ed.) (2002): *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press, pp. 348-367
- Kahneman, D. and Tversky, A. (1979): Prospect Theory: An Analysis of Decision under Risk. In: Kahneman, D. and Tversky, A. (ed.) (2000): *Choices, Values, and Frames*. Cambridge University Press, Pp. 17-43
- Kahneman, D. and Tversky, A. (1984): Choices, Values, and Frames. In: Kahneman, D. and Tversky, A. (ed.) (2000): *Choices, Values, and Frames*. Cambridge University Press, Pp.. 1-16
- Kahneman, D. and Tversky, A. (1986): Rational Choice and the Framing of Decisions. In: Kahneman, D. and Tversky, A. (ed.) (2000): *Choices, Values, and Frames*. Cambridge University Press, Pp. 209-223
- Kahneman, D. and Tversky, A. (2000): *Choices, Values, and Frames*. Cambridge University Press
- Kahneman, D.; Knetsch, J.L.; and Thaler, R.H. (1986a): Fairness and the Assumptions of Economics. *Journal of Business*, 59/4, pp. S285-S300
- Kahneman, D.; Knetsch, J.L.; and Thaler, R.H. (1986b): Fairness as a Constraint on Profit Seeking: Entitlements in the Market. In: Kahneman, D. and Tversky, A. (ed.) (2000): *Choices, Values, and Frames*. Cambridge University Press, Pp. 317-334
- Kahneman, D.; Slovic, P.; and Tversky, A. (ed.) (1982): *Judgment under Uncertainty: Heuristics and Biases*. Cambridge University Press
- Kahneman, D; Diener, E.; and Schwarz, N. (ed.) (1999): *Well-Being: The Foundations of Hedonic Psychology*. Russell Sage Foundation
- Kant, I. (1781/1787/2007): *Critique of Pure Reason*. Penguin Classics
- Kant, I. (1785/2002): *Groundwork for the Metaphysics of Morals*. Yale University Press
- Kant, I. (1788/2004): *Critique of Practical Reason*. Dover Philosophical Classics

- Kant, I. (1793/1794/2009): *Religion within the Bounds of Bare Reason*. Hackett Publishing Company
- Karathanos, P. Pettypool, M.D.; and Troutt, M.D. (1994): Sudden Lost Meaning. *Management Decision*, 32/1, pp. 15-19
- Katz, D. and Kahn, R. (1966/1978): *The Social Psychology of Organizations*. John Wiley and Sons
- Kaufmann, W. (ed.) (1973): *The Portable Nietzsche*. Penguin Books
- Keynes, J.M. (1936/2008): *The General Theory of Employment, Interest and Money*. Atlantic Publishers and distributors LTD
- Knight, J. and Ensminger, J. (1998): Conflict over Changing Social Norms: Bargaining, Ideology, and Enforcement. Brinton, M.C. and Nee, V. (1998): *The New Institutionalism in Sociology*. Stanford University Press, pp. 105-127
- Knoke, D.; Bohrnstedt, G.; and Mee, A. (2002): *Statistics for Social Data Analysis*. F. E. Peacock Publishers
- Korsgaard, C.M. (1986): Skepticism about Practical Reason. *The Journal of Philosophy*, 83/1, pp. 5-25
- Korsgaard, C.M. (1996): *Creating the Kingdom of Ends*. Cambridge University Press
- Korsgaard, C.M. (2008): *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. Oxford University Press
- Korsgaard, C.M. (2009): *Self-Constitution: Agency, Identity, and Integrity*. Oxford University Press
- Kosfeld, M.; Heinrichs, M.; Zak, P.J.; Fischbacher, U.; and Fehr, E. (2005): Oxytocin Increases Trust in Humans. *Nature*, 435, pp. 673-676
- Koudenburg, N.; Postmes, T.; and Gordijn, E. (2013): Resounding Silence: Subtle Norm Regulations in Everyday Interactions. *Social Psychology Quarterly*, 76/3, pp. 224-241
- Krosnick, J. and Presser, S. (2010): Question and Questionnaire Design. Pp. 263-313
- Kuhn, T. (1962/1996): *The Structure of Scientific Revolutions*. The University of Chicago Press
- Kuzon, W.; Urbanek, M.; and McCabe, S. (1996): The Seven Deadly Sins of Statistical Analysis. *Annals of Plastic Surgery*, 37/3, pp. 265-272
- Lakatos, I. (1970): Falsification and the Methodology of Scientific Research Programmes. In: Schick, T. (ed.) (1999): *Readings in the Philosophy of Science*. McGraw-Hill, pp. 170-196
- Lawler, E. E. (1994): From Job-Based to Competency-Based Organizations. *Journal of Organizational Behavior*, 15/1, pp. 3-15

- Lawler, E. E. III (1986b): What's Wrong with Point-Factor Job Evaluation. *Compensation and Benefits Review*, 18/5, pp. 20-28
- Lawler, E. E. III (2011): Creating a New Employment Deal: Total Rewards and the New Workforce. *Organizational Dynamics*, 40, pp. 302-309
- Lawler, E.E. (1971): *Pay and Organizational Effectiveness: A Psychological View*. McGraw-Hill
- Lawler, E.E. and Ledford, G. (1992): A Skill-based Approach to Human Resource Management. *European Management Journal*, 10/4, pp. 383-391
- Lawler, E.E. III (1986a): *High-Involvement Management*. Jossey-Bass Publishers
- Lawler, E.E. III (1990): *Strategic Pay*. Jossey-Bass Publishers
- Lawler, E.E. III (1992): *The Ultimate Advantage*. Jossey-Bass Publishers
- Lawler, E.E. III (2000): *Rewarding Excellence*. Jossey-Bass Publishers
- Lawler, E.E. III (2003): *Treat People Right!* Jossey-Bass Publishers
- Lawler, E.E. III (2008): *Talent*. Jossey-Bass Publishers
- Lawler, E.E. III and Worley, C.G. (2006) *Built to Change*. Jossey-Bass Publishers
- LeCompte, M. and Goetz, J. (1982): Problems of Reliability and Validity in Ethnographic Research. *Review of Educational Research*, 52/1, pp. 31-60
- Ledford, G.; Heneman, R.; and Salimäki, A. (2008): Skills, Knowledge, and Competency-based Pay. In: Berger, L. and Berger, D. (ed.) (2008): *The Compensation Handbook*. McGraw Hill, pp. 143-159
- Lee, C.; Law, K.S.; and Bobko, P. (1999): The Importance of Justice Perceptions on Pay Effectiveness: A Two-Year Study of a Skill-Based Pay Plan. *Journal of Management*, 25/6, pp. 851-873
- Lee, R.M. (1993): *Doing Research on Sensitive Topics*. Sage Publications
- Lemieux, T.; MacLeod, W.B.; and Parent, D. (2009): Performance Pay and Wage Inequality. *The Quarterly Journal of Economics*, CXXIV/1, pp. 1-49
- Leonard, N.H.; Beauvais, L.L.; and Scholl, R.W. (1999): Work Motivation: The Incorporation of Self-Concept-Based Theories. *Human Relations*, 52/8, pp. 969-998
- Lesses, G. (1987): Weakness, Reason, and the Divided Soul in Plato's Republic. *History of Philosophy Quarterly*, 4/2, 147-161
- Leventhal, G. (1980): What Should be Done with Equity Theory? In: Gergen, K.J.; Greenberg, M.; and Willis, R.H. (ed.) (1980): *Social Exchange*. Plenum Press, pp. 27-56
- Levine, D. (1991): Cohesiveness, Productivity, and Wage Dispersion. *Journal of Economic Behavior and Organization*, 15, pp. 237-255

- Lewis-Beck, M.; Bryman, A.; and Liao, T. (ed.) (2004): *Encyclopedia of Social Science Research Methods*. Sage Publications
- Liao, T. (2004): Odds Ratio. In: Lewis-Beck, M.; Bryman, A.; and Liao, T. (ed.) (2004): *Encyclopedia of Social Science Research Methods*. Sage Publications (entry: Odds ratio)
- Lieber, E. (2009): Mixing Qualitative and Quantitative Methods: Insights into Design and Analysis Issues. *Journal of Ethnographic and Qualitative Research*, 3, pp. 218-227
- Lind, E.A. (2001): Fairness Heuristic Theory: Justice Judgments as Pivotal Cognitions in Organizational Relations. In: Greenberg, J. and Cropanzano, R. (ed.) (2001): *Advances in Procedural Justice*. Stanford University Press, pp. 56-88
- Lysgaard, S. (1961/2001): *Arbeiderkollektivet: En Studie i de Underordnedes Sosiologi*. Universitetsforlaget
- Mahoney, T. (1989): Multiple Pay Contingencies: Strategic Design of Compensation. *Human Resource Management*, 28/3, pp. 337-347
- March, J.G. (2006): Rationality, Foolishness, and Adaptive Intelligence. *Strategic Management Journal*, 27, pp. 201-214
- March, J.G. and Olsen, J.P. (1975): The Uncertainty of the Past. *European Journal of Political Research*, 3, pp. 147-171
- March, J.G.; Sproull, L.S.; Tamuz, M. (1991): Learning from Samples of One or Fewer. *Organization Science*, 2/1, pp. 1-13
- Margetts, H.; 6, P.; and Hood, C. (ed.)(2010): *Paradoxes of Modernization: Unintended Consequences of Public Policy Reform*. Oxford University Press
- Marsden, D. (1999): *A Theory of Employment Systems: Micro-Foundations of Societal Diversity*. Oxford University Press
- Marsden, D. (2000): A Theory of Job Regulation, the Employment Relationship, and the Organization of Labour Institutions. *Industrielle Beziehungen*, 7/4, pp. 320-349
- Marsden, D. (2004a): The Role of Performance-Related Pay in Renegotiating the “Effort Bargain”: The Case of the British Public Service. *Industrial and Labor Relations Review*, 57/3, pp. 350-370
- Marsden, D. (2004b): The Network Economy and Models of the Employment Relation. *British Journal of Industrial Relations*, 42/4, pp. 659-684
- Marsden, D. (2010): The Paradox of Performance-Related Pay Systems: Why Do we Keep Adopting Them in the Face of Evidence that they Fail to Motivate? In: Margetts, H.; 6, P.; and Hood, C. (ed.)(2010): *Paradoxes of Modernization: Unintended Consequences of Public Policy Reform*. Oxford University Press

- Marsden, D. (2013): Individual Voice in Employment Relationships: A Comparison under Different Forms of Workplace Representation. *Industrial Relations*, 52/S1, pp. 221-258
- Marsden, P.M. and Wright, J.D. (ed.) (2010): *Handbook of Survey Research*. Emerald Group Publishing Ltd.
- Maslow, A. (1943): A Theory of Human Motivation. *Psychological Review*, 50, pp. 370-396
- Maslow, A. H. (1965/2011): *Toward a Psychology of Being*. Martino Publishing
- Maus, M. (1922/1966): *The Gift*. Cohen and West LTD
- Maxwell, J. and Mittapalli, K. (2010): Realism as a Stance for Mixed Methods Research. In: Tashakkori, A. and Teddlie, C. (ed.) (2010): *Sage Handbook of Mixed Methods in Social and Behavioral Research*. Sage Publications, pp. 145-167
- Mayo, E. (1949/2000): *The Social Problems of an Industrial Civilization*. Routledge
- McGregor, D. (1960/2006): *The Human Side of Enterprise*. McGraw-Hill
- Mead, G.H. (1934/1992): *Mind, Self, and Society*. The University of Chicago Press
- Medvec, V. H.; Madey, S.F.; and Gilovich, T. (1995): When Less is More: Counterfactual Thinking and Satisfaction among Olympic Medalists. In: Gilovich, T.; Griffin, D.W.; and Kahneman, D. (ed.) (2002): *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press, pp. 625-636
- Merton, R. K. (1957): The Role-Set: Problems in Sociological Theory. *The British Journal of Sociology*, 8/2, pp. 106-120
- Merton, R. K. (1968): *Social Theory and Social Structure*. The Free Press
- Merton, R.K. (1938/1968): Social Structure and Anomie. In: Merton, R. K. (1968): *Social Theory and Social Structure*. The Free Press
- Messick, D.M. and Sentis, K.P. (1979): Fairness and Preference. *Journal of Experimental Social Psychology*, 15, pp. 418-434
- Milkovich, G.; Newman, J.M.; and Gerhart, B. (2011): *Compensation*. McGraw Hill
- Mill, J. S. (1874/2000): *Essays on Some Unsettled Questions of Political Economy*. Batoche Books
- Mill, J.S. (1863/2001): *Utilitarianism*. Batoche Books
- Mill, J.S. (1882/2009): *A System of Logic*. Harper and Brothers, Project Gutenberg
- Montaigne, M. (1580/2003): *Michel De Montaigne – The Complete Essays*. Penguin Classics

- Moore, M.S. (2009): *Causation and Responsibility: An Essay in Law, Morals, and Metaphysics*. Oxford University Press
- Morgan, D. (1996): Focus Groups. *Annual Review of Sociology*, 22, pp. 129-152
- Morgan, D. (1998): Practical Strategies for Combining Qualitative and Quantitative Methods. *Qualitative Health Studies*, 8/3, pp. 362-376
- Morgan, D. (2007): Paradigms Lost and Pragmatism Regained: Methodological Implications of Combining Qualitative and Quantitative Methods. *Journal of Mixed Methods Research*, 1/1, pp. 48-76
- Nee, V. and Ingram, P. (1998): Embeddedness and Beyond: Institutions, Exchange, and Social Structure. In: Brinton, M.C. and Nee, V. (1998): *The New Institutionalism in Sociology*. Stanford University Press, pp. 19-46
- Nichols, S. (2004): *Sentimental Rules*. Oxford University Press
- Nicklin, J. (2013): Expertise, Counterfactual Thinking, and Fairness Perceptions: A Test of Fairness Theory. *Social Justice Research*, 26, pp. 42-60
- Nicklin, J.M.; Greenbaum, R.; McNall, L.A.; Folger, R.; and Williams, K.J. (2011): The Importance of Contextual Variables when Judging Fairness: An Examination of Counterfactual Thoughts and Fairness Theory. *Organizational Behavior and Human Decision Processes*, 114, pp. 127-141
- Nietzsche, F. (1873/1976): On Truth and Lie in an Extra-Moral Sense. In: Kaufmann, W. (ed.) (1973): *The Portable Nietzsche*. Penguin Books, pp. 42-50
- Nietzsche, F. (1887/2000): On the Genealogy of Morals. In: Nietzsche, F. (2000): *Basic Writings of Nietzsche*. The Modern Library Classics, pp. 437-600
- Nietzsche, F. (2000): *Basic Writings of Nietzsche*. The Modern Library Classics
- Niglas, K. (2001): How the Novice Researcher can make Sense of Mixed Methods Design. *International Journal of Multiple Research Approaches*, 3/1, pp. 34-46
- Non, A. (2012): Gift-exchange, Incentives, and Heterogeneous Workers. *Games and Economic Behavior*, 75, pp. 319-336
- Nozick, R. (1993): *The Nature of Rationality*. Princeton University Press
- Opp, K-D. (2001): Social Networks and the Emergence of Protest Norms. In: Hechter, M. and Opp, K-D. (ed.) (2001): *Social Norms*. Russell Sage Foundation, pp. 234-273
- Organ, D. W. (1997): Organizational Citizenship Behavior: It's Construct Clean-Up Time. *Human Performance*, 10/2, pp. 85-97
- Ouchi, W. and Price, R. (1978/1993): Hierarchies, Clans, and Theory Z. *Organizational Dynamics*, 21/4, pp. 62-70
- Pearl, J. (2009): *Causality*. Cambridge University Press

- Perry, J.L. and Wise, L.R. (1990): The Motivational Bases of Public Service. *Public Administration Review*, 50/3, pp. 367-373
- Peters, T. and Waterman, R. (1982/2004): *In Search of Excellence*. Profile Books
- Pfeffer, J. (1994): *Competitive Advantage Through People*. Harvard University Press
- Pfeffer, J. (1998): *The Human Equation*. Harvard Business Review Press
- Pfeffer, J. and Langdon, N. (1993): The Effect of Wage Dispersion on Satisfaction, Productivity, and Working Collaboratively – Evidence from College and University Faculty. *Administrative Science Quarterly*, 38/3, pp. 382-407
- Pickett, C.; Gardner, W.; and Knowles, M. (2004): Getting a Cue: The Need to Belong and Enhanced Sensitivity to Social Cues. *Personality and Social Psychology Bulletin*, 30, pp. 1095-1107
- Pinker, S. (2011): *The Better Angels of our Nature*. Allen Lane
- Plato (360BC/1961a): Phaedo. In: Hamilton, E. and Cairns, H. (ed.) (1961): *Plato – The Collected Dialogues*. Princeton University, pp. 40-99
- Plato (360BC/1961b): Philebus. In: Hamilton, E. and Cairns, H. (ed.) (1961): *Plato – The Collected Dialogues*. Princeton University, pp. 1086-1150
- Plato (360BC/1961c): Timaeus. In: Hamilton, E. and Cairns, H. (ed.) (1961): *Plato – The Collected Dialogues*. Princeton University, pp. 1151-1212
- Plato (380 BC/1961a): Republic. In: Hamilton, E. and Cairns, H. (ed.) (1961): *Plato – The Collected Dialogues*. Princeton University, pp. 575-845
- Plato (380BC/1961b): Protagoras. In: Hamilton, E. and Cairns, H. (ed.) (1961): *Plato – The Collected Dialogues*. Princeton University, pp. 308-353
- Popper, K. R. (1935/2002): *The Logic of Scientific Discovery*. Routledge
- Popper, K. R. (1957/2002): *The Poverty of Historicism*. Routledge
- Popper, K. R. (1963/2002): *Conjectures and Refutations*. Routledge
- Porter, L. and Lawler, E.E. (1968): *Managerial Attitudes and Performance*. Richard D. Irwin, INC.
- Prahalad, C.K. and Hamel, G. (1990): The Core Competence of the Corporation. *Harvard Business Review*, May-June, pp. 79-90
- Prinz, J. (2012): *Beyond Human Nature*. Allen Lane
- Rabin, M. (1993): Incorporating Fairness into Game Theory and Economics. *The American Economic Review*, 83/5, 1281-1302
- Rescher, N. (2002): *Fairness*. Transaction Publishers

- Robinson, S.L.; O'Reilly, J.; and Wang, Wei (2013): Invisible at Work: An Integrated Model of Workplace Ostracism. *Journal of Management*, 39/1, pp. 203-231
- Roethlisberger, F.J. and Dickson, W.J. (1939/1975): *Management and the Worker*. Harvard University Press
- Rosen, G. (2011): Causation, Counterfactual Dependence and Culpability. *Rutgers Law Journal*, 42/2, pp. 405-437
- Rossi, P. and Anderson, A. (1982): The Factorial Survey Approach. In: Rossi, P. and Nock, S. (ed.) (1982): *Measuring Social Judgments*. Sage Publications, pp. 15-69
- Rossi, P. and Nock, S. (ed.) (1982): *Measuring Social Judgments*. Sage Publications
- Rousseau, D.M. (1995): *Psychological Contracts in Organizations*. Sage Publications
- Rousseau, J.J. (1755/1984): *A Discourse on Inequality*. Penguin Classics
- Ryan, R. and Deci, E. (2006): Self-Regulation and the Problem of Human Autonomy: Does Psychology Need Choice, Self-Determination, and Will? *Journal of Personality*, 74/6, pp. 1557-1586
- Scheler, M. (1915/1994): *Ressentiment*. Marquette University Press
- Scheler, M. (1915/2004): *Das Ressentiment im Aufbau der Moralen*. Klostermann Seminar
- Schick, T. (ed.) (1999): *Readings in the Philosophy of Science*. McGraw-Hill
- Schminke, M. (ed.) (1998): *Managerial Ethics: Moral Management of People and Processes*. Lawrence Erlbaum, Inc.
- Schuster, J.R. and Zingheim, P.K. (1992): *The New Pay*. Jossey-Bass Publishers
- Schuster, J.R. and Zingheim, P.K. (2000): *Pay People Right*. Jossey-Bass Publishers
- Schwarz, N. and Vaughn, L.A. (2002): The Availability Heuristic Revisited: Ease of Recall and Content of Recall as Distinct Sources of Information. In: Gilovich, T.; Griffin, D.W.; and Kahneman, D. (ed.) (2002): *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press, pp. 103-120
- Shadish, W.; Cook, T.; and Campbell, D. (2002): *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Houghton Mifflin Company
- Shamir, B. (1990): Calculations, Values, and Identities: The Sources of Collectivistic Work Motivation. *Human Relations*, 43/4, pp. 313-332
- Simon, H. (1945/1997): *Administrative Behavior*. The Free Press
- Simon, H. (1951): A Formal Theory of the Employment Relationship. *Econometrica*, 19/3, pp. 293-305
- Simon, H. (1983): *Reason in Human Affairs*. Stanford University Press

- Simon, H. (1991): Organizations and Markets. *Journal of Economic Perspectives*, 5/2, pp. 25-44
- Singer, T.; Seymour, B.; O'Doherty, J.; Stephan, K.; Dolan, R.; and Frith, C. (2006): Empathetic Neural Responses are Modulated by the Perceived Fairness of Others. *Nature*, 439, pp. 466-469
- Smelser, N.J. and Swedberg, R. (ed.) (2005): *The Handbook of Economic Sociology*. Russell Sage Foundation
- Smith, A. (1759/2009): *The Theory of Moral Sentiments*. Penguin Classics
- Smith, A. (1776/2003): *The Wealth of Nations*. Bantam Books
- Smith, D. (1923/1951): *History of Mathematics*. Dover Publications
- Sperber, D. (2011): A Naturalist Ontology for Mechanistic Explanation in the Social Sciences. Pp. 64-78
- Spinoza, B. (1677/1996): *Ethics*. Penguin Classics
- Tashakkori, A. and Teddlie, C. (ed.) (2003): *Handbook of Mixed Methods in Social and Behavioral Research*. Sage Publications
- Tashakkori, A. and Teddlie, C. (ed.) (2010): *Sage Handbook of Mixed Methods in Social and Behavioral Research*. Sage Publications
- Taylor, F.W. (1911/1998): *The Principles of Scientific Management*. Dover Publications
- Thaler, R.H. (1980): Towards a Positive Theory of Consumer Choice. In: Kahneman, D. and Tversky, A. (ed.) (2000): *Choices, Values, and Frames*. Cambridge University Press, Pp. 269-288
- Thibaut, J. and Walker, L. (1978): A Theory of Procedure. *California Law Review*, 66/3, pp. 541-566
- Tourangeau, R. and Bradburn, N. (2010): The Psychology of Survey Response. Pp. 315-346
- Twenge, J.M.; Baumeister, R.; Tice, D.M.; and Stucke, T.S. (2001): If you Can't Beat Them, Join Them: Effects of Social Exclusion on Aggressive Behavior. *Journal of Personality and Social Psychology*, 81/6, pp. 1058-1069
- Tyler, T. and Blader, S. (2003): *The Group Engagement Model: Procedural Justice, Social Identity, and Cooperative Behavior*. 7/4, pp. 349-361
- Tyler, T.R. and Lind, E.A. (1992): A Relational Model of Authority in Groups. In: Zanna, M.P. (ed.) (1992): *Advances in Experimental Social Psychology*. Academic Press, 25, pp. 115-191
- Ulrich, D.; Younger, J.; and Brockbank, W. (2008): The Twenty-first-century HR Organization. *Human Resource Management*, 47/4, pp. 829-850

- Vohs, K.; Mead, N.; and Goode, M. (2006): The Psychological Consequences of Money. *Science*, 314, pp. 1154-1156
- Vroom, V.H. (1964/1995): *Work and Motivation*. Jossey-Bass Publishers
- Weber, M. (1922/1968): *Economy and Society*. Two vols., University of California Press
- White, M.; Hill, S.; McGovern, P.; Mills, C.; and Smeaton, D. (2003): High performance Management Practices, Working Hours, and Work-Life Balance. *British Journal of Industrial Relations*, 41/2, pp. 175-195
- White, R.W. (1959): Motivation Reconsidered. *Psychological Review*, 66/5, 297-333
- Williamson, O. (1996): *The Mechanisms of Governance*. Oxford University Press
- Wilson, T.B. and Malanowski, S. (2008): Performance Management Best Practices. In: Berger, L. and Berger, D. (ed.) (2008): *The Compensation Handbook*. McGraw Hill, pp. 447-459
- Wolf, S. (2013): The Moral of Moral Luck. *Philosophic Exchange*, 31/1, article 1.  
Available at  
[http://digitalcommons.brockport.edu/cgi/viewcontent.cgi?article=1052&context=phil\\_ex](http://digitalcommons.brockport.edu/cgi/viewcontent.cgi?article=1052&context=phil_ex)
- Zadro, L.; Williams, K.; and Richardson, R. (2004): How Low can you Go? Ostracism by a Computer is Sufficient to Lower Self-reported Levels of Belong, Control, Self-esteem, and Meaningful Existence. *Journal of Experimental Social Psychology*, 40, pp. 560-567

## Appendix-1: Translated-Edition of the questionnaire

Peter and Søren are both employed as project-consultants in the same organization. Søren has been employed in the organization for five years, while Peter has only been employed for one year. They both have a cand.merc. Peter's monthly salary is 34,064 Dkr., while Søren's monthly salary is 49,064 Dkr. They are both viewed as being competent to their job, just as they both supply an effort which is above the expected level. In addition, they share an office. In the previous period Peter was, after hard work, so lucky as to become part of a strategically important project. Søren also tried to become a part of this project, but did not succeed. The project turned out to be a success and each member of the group received a one-off bonus on 15,000 Dkr. This entailed that Peter's salary last period was 49,064 Dkr.

1. Please evaluate whether you find it alright that Peter's salary plus bonus in the previous period was equal to Søren's salary?
2. Please evaluate whether you find the foundation of why Peter's salary with the bonus was equal to Søren's salary alright?
3. Please evaluate whether you find Peter's active outreach to the project team to be a violation of some unwritten rule?
4. How significant do you find the bonus to be for Peter?
5. How significant do you find Peter's bonus to be for Søren?
6. How do you evaluate the equality will influence the relation between Peter and Søren?

Jeanette and Pernille are both employed in the same organization. Jeanette was hired 4 years ago, while Pernille was hired 2 years later. Both employees hold a HD, just like they are both viewed as experts in their field. Jeanette's average monthly salary is 39,550 Dkr., while Pernille's average monthly salary is 36,550 Dkr. Their job includes that the two must work together on the preparation of the department's budgets and other administrative tasks. Pernille have in the current period attended an on-the-job training course on a new IT-system and this has qualified her to be coordinator when the system is to be implemented next fall. Pernille choose to accept the company's offer because she perceived as a career opportunity. In relation to her attendance of the course, Pernille has received a monthly qualification allowance on 9,137 Dkr., this entails that her monthly salary now is 45,687 Dkr. Jeanette was also offered to attend the course but refused to do so.

7. Please evaluate whether you find the pay difference between Jeanette and Pernille alright?
8. Please evaluate whether you find the foundation of the pay difference between Jeanette and Pernille to be alright?

9. Please evaluate whether you find Pernille's acceptance of the opportunity offered to her to be a violation of some unwritten rule?
10. How significant do you find the quality allowance to be for Pernille?
11. How significant do you find Pernille's allowance to be for Jeanette?
12. How do you evaluate the difference will influence the relation between Pernille and Jeanette?

Jørgen and Anders are both employed as engineers in the same organization. They were both hired 5 years ago to aid the startup of a new department. Since their hiring, both Jørgen and Anders have worked on creating a well-functioning department. Both Jørgen and Anders are recognized as supplying a satisfactory performance. Jørgen is perceived as an expert in his field and very theoretically gifted, thus Anders often come to him for advice. Anders, on the other hand, is very pragmatic and has primarily been concerned with the management tasks. At the last performance appraisal, Jørgen was highly recognized for his theoretical knowhow. Anders was also positively recognized for his management skills and he was offered a take a MBA (Master in Business Administration – a relative expensive education).

13. Please evaluate whether you find it alright that Anders is offered the educational opportunity?
14. Please evaluate whether you find the foundation of Anders' offer alright?
15. Please evaluate whether you find that Anders is violation of some unwritten rule by being responsible for leadership related tasks?
16. How significant do you find the offer to be for Anders?
17. How significant do you find Anders' offer to be for Jørgen?
18. How do you evaluate the difference will influence the relation between Anders and Jørgen?

Introduction: Finally, I want to ask you some questions about yourself and your organization.

Gender?

Please state your age?

At which education are you currently enrolled?

Which HD 2 direction have you chosen or will you choose?

Please state your years of professional experience?

What is your most recently completed education?

Are you currently employed?

If you have answered No to the previous question there are no further questions. If you have answered Yes, please proceed.

What is your contractually agreed number of work hours?

What is the primary nature of your work?

Do you have any subordinates?

If Yes, please state how many?

Please state the pay forms you believe best to describe yours?

If you have a variable pay component attached to your pay please describe it?

Please state within which industries your organization has its main business areas?

## Appendix-2: Cover Letters

Email/letter to the companies: Original version

Jeg er i gang med at planlægge et forskningsprojekt i forbindelse med mit ph.d.-forløb ved Institut for Ledelse og Virksomhedsstrategi, Syddansk Universitet, og søger i den forbindelse virksomheder, der ønsker at deltage i min undersøgelse. Projektet omhandler relationen mellem løndifferentiering og sociale relationer mellem kollegaer, en længere projektbeskrivelse er vedlagt.

Selve undersøgelsen skal gennemføres i efteråret 2011 og skal selvfølgelig planlægges på en sådan måde, at den ikke forstyrrer dagligdagen for de involverede ledere og medarbejdere. Endvidere drejer det sig kun om nogle få interviews (ca. 10) og en spørgeskemaundersøgelse.

De indsamlede data og analyser vil blive publiceret i en rapport og i andre relevante sammenhænge. I den forbindelse vil alle medarbejderne være anonyme, ligesom virksomhedens navn heller ikke vil fremgå af den endelige rapport eller kildematerialet.

Som en opfølgning på undersøgelsen vil der være mulighed for at få tilsendt den del af rapporten, der omhandler organisationen. Ligesom jeg gerne kommer ud og præsenterer resultaterne af undersøgelsen.

Jeg håber dette har vakt jeres interesse og at I ønsker at deltage og skulle I have behov for yderligere information er I velkomne til at kontakte mig. Såfremt I ønsker at deltage bedes I kontakte mig senest d. 23. februar 2011.

Med venlig hilsen,

Martin Lund Petersen, Cand.merc.  
Ph.d.-studerende  
Syddansk Universitet, Campus Slagelse  
Mobil tlf. 3190 6754

Translated version:

I am currently planning a research project in relation with my PhD-work at University of Southern Denmark, Department of Leadership and Corporate strategy. In relation with this, I seek some companies who would be interested in participating in my research. The focus of the project is on the relation between pay differentiation and social relations between colleagues.

The research will be carried out in the fall of 2011 and will, of course, be planned in a manner which minimizes the disturbance of the parties involved – leaders as well as employees. In addition, I have only planned to carry out, approximately, 10 interviews at each company.

The collected data and analysis hereof will be published in a final report. All the participating parties will be anonymous, just as the company's name will not be mentioned in the in the project report.

Following the research process, the participating companies will have the opportunity of obtaining the part of the research which focuses on their company, just as I will be more than happy to present the results.

I hope that you will like to participate and should you need further information, you are more than welcome to contact me.

Email to students:

Original versions:

Kære HD-studerende,

Gennem de sidste 20 år er der sket store ændringer i forholdet mellem organisationen og medarbejderen, bl.a. er det centralt for at forstå moderne arbejdsliv, at man inddrager begreber som flextid, hjemmearbejde, selvledelse, performance evaluering, lønsamtaler, udviklingssamtaler, bonusordninger og individualiseret løndannelse, hvor mere eller mindre automatisk lønregulering og anciennitet er blevet erstattet af individuelle løntil-læg og resultatløn.

Som det fremgår, har løn indtaget en central plads som et aktivt ledelsesværktøj til bl.a. motivation af medarbejderen, til at skabe mållighed, til at skabe lyst til selvudvikling mv. Dette skal dog også ses i lyset af, at individualiseret løndannelse bl.a. har åbnet op for muligheden for at skabe lønforskelle mellem kollegaer, der har næsten samme job. Tidligere studier har primært taget fat på relationen mellem organisationen og medarbejderen og ikke direkte set på konsekvenserne af sådanne lønforskelle for professionelle relationer i organisationer.

Formålet med den undersøgelse som jeg vil bede dig bidrage til er derfor netop at belyse løns påvirkning af disse kollegiale relationer for at forstå, hvordan løn virker i organisationer. Jeg håber derfor, at du vil være med til at skabe viden om netop dette vigtige område ved at udfylde spørgeskemaet.

Undersøgelsen er anonym og din besvarelse vil blive behandlet med fortrolighed. Dit navn figurerer kun i mail-listen og vil ikke blive knyttet til din besvarelse. Denne ano-

nymisering af data vil blive overholdt også i det videre arbejde med data og i den endelige rapport.

Undersøgelsen udgør datagrundlaget for mit ph.d.-projekt ved Syddansk Universitet, Institut for Ledelse og Virksomhedsstrategi. Hovedvejleder på projektet er professor, dr. merc. Steen Scheuer. Resultaterne af undersøgelsen vil blive fremlagt i en endelig rapport, men der er mulighed for, at du kan få tilsendt et resume af undersøgelsens resultater således, at du kan se den generelle holdning på HD til sammenhængen mellem løn og kollegiale relationer. Hvis du ønsker at få tilsendt et resume skal du blot kontakte mig på e-mail.

Din besvarelse er meget vigtig for undersøgelsen, hvorfor jeg håber at du vil udfylde spørgeskemaet. Du finder spørgeskemaet ved at trykke på nedenstående link. Når spørgeskemaet er åbnet skal du blot følge instruktionerne.

Skulle du have nogle spørgsmål er du meget velkommen til at kontakte mig på telefon eller e-mail.

Follow-up Email:

Kære HD-studerende,

For nogle uger siden fremsendte jeg et spørgeskema, der havde til formål at belyse forholdet mellem løn og kollegiale relationer i organisationer. Hvis du allerede har svaret på spørgeskemaet takker jeg meget for din deltagelse. Hvis ikke du har svaret på spørgeskemaet endnu bedes du gøre det i dag.

Formålet med den undersøgelse jeg vil bede dig bidrage til er at belyse løns påvirkning af kollegiale relationer i organisationer – et område som tidligere undersøgelser ikke har fokuseret på som centralt for at forstå, hvordan løn virker i organisationer. Jeg håber derfor, at du vil være med til at skabe viden om netop dette vigtige område ved at udfylde spørgeskemaet.

Endvidere vil jeg gerne gøre opmærksom på, at undersøgelsen er anonym og at din besvarelse vil blive behandlet med fortrolighed.

Undersøgelsen udgør datagrundlaget for mit ph.d.-projekt ved Syddansk Universitet, Institut for Ledelse og Virksomhedsstrategi. Hovedvejleder på projektet er professor, dr. merc. Steen Scheuer. Resultaterne af undersøgelsen vil blive fremlagt i en endelig rapport, men der er mulighed for, at du kan få tilsendt et resume af undersøgelsens resultater således, at du kan se den generelle holdning på HD til sammenhængen mellem løn og kollegiale relationer. Hvis du ønsker at få tilsendt et resume skal du blot kontakte mig på e-mail.

Din besvarelse er meget vigtig for undersøgelsen, hvorfor jeg håber at du vil udfylde spørgeskemaet. Du finder spørgeskemaet ved at trykke på nedenstående link. Når spørgeskemaet er åbnet skal du blot følge instruktionerne.

Skulle du have nogle spørgsmål er du meget velkommen til at kontakte mig på telefon eller e-mail.

Translated versions:

First email:

Dear HD-Students

Through the last 20 years a lot have happened with the relation between the organization and the employee. For example, in order to understand the modern employment relation, one has to take concepts like flex-hours, performance evaluation, employee development conversations, bonus pay, and individualized compensation into account, just as one also have to take into account that matters previously related with automatic regulation of pay and seniority have been, more or less, replaced by individual bonuses and performance pay.

Pay plays, in other words, a more central role as a modern governance instrument of, among other things, motivation, to create goal congruence, to produce motivation to self-development, etc. This should also be viewed in the light of the fact that individualized pay has generated the possibility for pay differences between employees holding similar jobs. Previous studies have primarily been concerned with the relation between the organization and its employees and have not, as such, focused on the potential consequences of such pay differences on the relation between employees.

The purpose of this study, which I will ask you to contribute to, is therefore to provide some insight into the influence of pay on these relations in order to understand how pay functions. Hence, I hope that you will take part in the production of knowledge on this important matter by participating in the survey.

The research is anonymous and your response will be treated in confidence. Your name will only be present in the head-line of the email and will not be tied to your response. This will also apply in the final report.

The survey provides the data to my PhD-project at University of Southern Denmark, Department of Leadership and Corporate Strategy. Supervisor on the project is Steen Scheuer. The results of the research will be published in a final report, but you will have the opportunity of obtaining a resume.

Your participation in this survey is very important, thus I hope that will participate.

Follow-up email:

A couple of weeks ago I asked you to participate in my research on the relation between pay and relations between employees. If you already have participated I would like to thank you. If you have not yet participated I will ask you please to do it today.

The purpose of this study, which I will ask you to contribute to, is therefore to provide some insight into the influence of pay on these relations in order to understand how pay functions. Hence, I hope that you will take part in the production of knowledge on this important matter by participating in the survey.

The research is anonymous and your response will be treated in confidence. Your name will only be present in the head-line of the email and will not be tied to your response. This will also apply in the final report.

The survey provides the data to my PhD-project at University of Southern Denmark, Department of Leadership and Corporate Strategy. Supervisor on the project is Steen Scheuer. The results of the research will be published in a final report, but you will have the opportunity of obtaining a resume.

Your participation in this survey is very important, thus I hope that will participate.

## Appendix-3: Interview Summaries

### Organization X

#### *Case 1*

The respondent argues that the reward difference is alright, because the receiving individual did something to obtain it. Of course, it is not fun for the non-receiving individual, but the receiving agent did not do anything wrong. There is not, as such, any rules preventing the receiving individual from pursuing his ambitions as long as he does so in an appropriate manner.

#### *Case 2*

Here the respondent emphasizes that the non-receiving agent only has herself to blame, in the sense that she could have accepted the offer. Even though, the non-receiving agent only has herself to blame, the respondent does not believe that is easier to accept. On the contrary, she argues, it might be easier to accept if someone else had made the decision, because then she would not blame herself. Furthermore, the respondent argues that this might not affect the relation between the receiving and non-receiving negatively, insofar as both parties need to behave in an appropriate manner.

#### *Case 3*

Again the respondent emphasizes that the receiving agent cannot be blamed for bringing about the new outcome situation, just as she emphasizes that the non-receiving agent might not be interested in participating in such an education. In continuation of the previous argument on acceptance, the respondent argues that in this case, in juxtaposition to case 2, it might be easier to accept the difference because the non-receiving agent cannot blame himself – someone else made the decision.

## **Organization Y**

### *Case 1*

The respondents here emphasized the fact that the receiving agent did something to obtain the reward – he did something besides his work. The respondents argue that the non-receiving might feel bad not obtaining a reward, but this emotional experience is not the result of the previous relational status quo; rather the difference might matter because it is perceived as a loss of appreciation – not in relational, but temporal terms. Furthermore, the respondents also argue that the initial pay difference between the receiving and non-receiving individuals might be too high – it is not however unfair that the individual with the highest seniority obtains the highest pay – only the size of the difference.

### *Case 2*

Here the respondents emphasize that the pay difference is fair, because the receiving individual did something – as one of respondents argue: nobody does that solely for the fun of it. In continuation of their previous argument on appreciation, the respondents argue that if the relational difference between before and after matters it is solely because it is felt as a loss of appreciation.

### *Case 3*

Here the respondents emphasize that they find it fair that the receiving individual was offered the educational opportunity, because the non-receiving agent might not have the necessary competences needed – it is, then, fair to differentiate based on competences, at least, in matters related with educational opportunities. Furthermore, the respondents also argue that the non-receiving perhaps should try to find an on-the-job-course that he would like to attend.