Trust as a Management Technology: The Instrumentalization of Trust

Martin Lund Petersen, PhD (mape@sdu.dk)
Department of Leadership and Corporate Strategy, University of Southern Denmark

Kristian Rune Hansen, PhD-fellow (kruneh@sdu.dk)
Department of Leadership and Corporate Strategy, University of Southern Denmark

Abstract

This article is centered on the relationship between management as an instrumentalized practice, trust, reciprocity and the employment contract. Utilizing game theory, the article is focused on expanding the framework for analyzing and understanding trust-based employment relationships, as related to the conceptualization of trust as a rational/instrumental model and trust as socially embedded. The article is focused on integrating theoretical perspectives on trust with game theory conceptualization of employment contract and the employment relationship more generally. The article concludes on the social nature of the concept of trust, in a game theory perspective; among other things the importance of trust, when shaping the social foundation of the employment relation.

Introduction

Trust, as a theoretical concept, has been discussed in a wide range of academic traditions; from philosophy to sociology and psychology. In recent years, there has been an increasing interest in the concept of trust in the management field and especially in the research done on organizations and personnel management.

In the leadership, management and business fields, trust can - with a rough approximation - be said to have been discussed in three main overall categories of research. 1) Trust between organizations (*inter-organizational*), typically in the form of trust between co-operating organizations, for example supply chain partners or similar. 2) Trust between organizations and customers (*extra-organizational*), typically in a marketing perspective and finally, the focus of this paper, 3) trust between employers and employees, as well as trust between co-workers (*intra-organizational*), typically in a HRM or personnel management oriented perspective. In the most general sense, trust can be considered a fundamental requirement for personnel management in general and it is of singular importance when it comes to understanding the employee-employer/manager relationship, as well as employee-employee relationships and cooperation.

Categorizations and Definitions

The research undertaken in relation to the concept of trust is quite varied and expansive, and the concept of trust itself, has been defined and measured in numerous ways. As a method to categorize different forms of trust Lewicki and Bunker (1995), based on a distinction by Worshel (1979), suggest that rather than understanding trust as unidimensional, three forms of trust should be considered; intrapersonal, interpersonal and institutional trust respectively. These are understood to be separate but linked types of trust, which can change and interact depending on the relationship being examined.

The most often used definition of trust in a management perspective, is from Rousseau, Sitkin, Burt and Camerer (1998) who, building on a general consensus among researchers in many different fields, suggest a core trait of the way trust has been understood theoretically implies a willingness to be vulnerable. Based on this, they define trust in the following way: "*a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another*" (Rousseau, Sitkin, Burt, & Camerer, 1998, p. 395).

This definition is tied to a rational choice or instrumental conceptualization of trust, that is, predictability as it relates to expected outcomes of a future behaviour from others, is centred on the interpersonal and rational perspective and as Tyler (2003) suggests, the rational or instrumentality based definition becomes is inadequate, especially in situations where interpersonal predictability is low, but trust is high regardless. To supplement this rational/instrumental conceptualization of trust, Tyler proposes the concept of *social trust* as an overarching category, and a way to understand trust more completely, and likens the distinction to concepts from the organizational justice literature; instrumental and procedural justice, but with motive-based trust as form of social trust, introduced as a separate and distinct concept from procedural justice.

Procedural justice and institutional, or systemic, forms of trust and can, in some situations, be understood as a substitute for trust, when it comes to some of the problems with regards to trust-relations. In other words, while procedural justice is not directly tied to the manager-employee relation, it can in daily work-place practice be important when it comes to various trust-related issues, while the manager himself, and by extension hereof - the relationship between employee and manager - is perhaps less important. That is, it's possible that if an employee feels that there are reasonable structural/systemic procedures in place, it can be considered easier to cope with an untrustworthy manager (and vice versa). The genesis of these structural trust-patterns is not necessarily organizational solely, as they can be possibly be enforced both a societal scale or locally, such as labour laws or trade union agreements, or in terms of the organization itself, with transparent

procedures regarding the hiring and firing of employees, promotions, and more. However as shown by Tyler (2003) the effect of procedural justice and social-trust are empirically distinctive from one another. In this perspective, social motives, such as motive-based trust and procedural justice are considered internal, and actor-based, because they exist individually and separately from types of sanctions and incentives prevalent in a given organization, but are still considered separate and different from the rational/instrumental types of trust (Tyler, 2003, p. 559-560). It is further suggested that, motive-based trust consists of two primary elements; 1) shared background and values. 2) An understanding of why the other person is doing what he is doing (which is separate from the instrumental trust or the ability to predict how a person will react in the future). Motive-based trust is empirically distinct from procedural justice and it is shown that motive-based trust influences attitudes and extra-role behaviour, where procedural justice influences values and deference (Tyler, 2003, p. 564).

In a fairly recent article on interpersonal trust, Lewicki, Tomlinson and Gillespie (2006), categorize the existing research into two main conceptual groups, a behavioural and a psychological approach, with the psychological group being further divided into three main categories. The behavioural approach is centred in a rational choice perspective, and is defined in relation to willingness to cooperate and examined primarily in game theory terms (prisoners' dilemma and the variants relating to this). Later in this paper, we will return to this point, and attempt to introduce a social perspective on trust, in game theoretical terms. In the psychological approaches, three main variants are identified; a unidimensional, a two-dimensional and a transformational. In the unidimensinoal perspective, trust is understood to be a scale that goes from high trust at one end to high distrust at the other. The two-dimensional view, suggests two different scales, one for trust and one for distrust, and finally in the transformational perspective, in which trust is seen as resting on the basis of several different indicators, such as expected cost and benefits, knowledge, shared values, identities and more (Lewicki, Tomlinson & Gillespie, 2006, p. 994). The third variant of the psychological approach is further differentiated into three main levels of trust, deterrence-based, knowledge-based and identification-based trust.

Instrumentalization, incentivizing performance and management practice

Generally trust-based management is often juxtaposed with control-based management, with the latter typically considered somewhat antiquated, despite it the apparent dominance in most workplaces (Tyler 2003). Which we will return to later in this paper, where we also will argue that the trust based contract is characterized by the absence of control or inducement in compared to the performance contract.

HRM, or personnel, related management practice, that is the practice of doing management of people, is in general inherently instrumental, especially when it comes to the various aspects of management theory aimed at motivating employees to increased performance, such as incentivized/performance-based pay and the usage and description of other motivational tools. In the most basic sense, this is the case because these types of management theory, as practice, are utilized to attain a goal or objective of some sort, whether it is increased performance, alignment of interests, employee retention, reduction of employee absence or many of the other 'classics' in the HRM literature.

Theories on trust-based leadership and trust-based management, along with theories emphasizing the strategic role of HRM as facilitating the social bonding process between peers are all describing how the phenomenon of trust should be utilized in the knowledge-intensive organization employing highly specialized and talented employees. From a management perspective, trust can be viewed as a technology in the sense that it facilitates governance as a mixture of control, self-control, and trust while maintaining work motivation. The manager should, then, interact with her employees in a manner which symbolizes mutual trust – that is, the employees can trust her just as she trusts them to make the right decisions as we will also argue is the case with trust based contracts and their belief management function. Trust can, in other words, be utilized as an important means to facilitate the functioning of asymmetric relations (manager-employee), but it can also be viewed as an important means for facilitating knowledge sharing in symmetric

relations between employees. The question here is not whether trust is important in social relations asymmetric as well as symmetric, rather the question is whether this utilization of trust as a means to some particular end besides the continuation of social relations, i.e. the facilitation of organizational effectiveness? In this perspective, the central questions then becomes, is the instrumentalization of trust not diluting the meaning of trust and from a critical-ethical perspective, is governance through trust not just another way of *capitalizing* on human nature, like certain theories on employee loyalty? (Tepper 2000, Vigoda-Gadot 2006). The focus on facilitating co-operation and underlines the importance of trust in many workplaces, especially when it comes to knowledge intensive organizations. However, the importance of trust in relation to co-operation, promoting employee commitment and the various other positive effects of work-place trust is not contradictory, but rather adjacent to, the instrumentalization of trust as a management technology, or perhaps more accurately a technology of power, in a similar vein to Taylorism, as suggested from a theory of power perspective by Clegg, Courpasson & Phillips (2006, p. 26), and labour process theory by Knight and Willmott (1989) and expanded upon by Townley (1999). From this perspective, trust as a management technology in some sense represents the (foucaldian) shift from controlling the body of the worker, to controlling the soul, "*(…) legitimized by the uncontestable discourse of efficiency*" (Clegg, Courpasson & Phillips, 2006, p. 40).

Understanding various forms of management practice, as a technology of control is similar to the industrial relations perspective, where the focus is often on the labour control aspects of management technologies, and trust typically understood as fundamentally relational. It is suggested that strong and coercive management technologies and managerial control are reciprocated with low trust employee attitudes and behaviour (Watson, 1995, p. 292). Labour control is understood to involve three main components, direction, surveillance and discipline, and is introduced on the basis of attempting to overcome the so-called principal-agent problem, where employees' commitment to organizational/managerial goals cannot be taken for granted. However, the industrial relations perspective also highlights that there is a fundamental contradiction in

implementing (coercive) labour control, as is likely to destroy initiative, diligence and strong organizational commitment in the employee (Watson, 1995, p. 292). As such, the concept of trust comes at the forefront when examining employee-manager relations, in terms of control and surveillance.

Sewell and Barker (2006), exploring the concept of surveillance and control in a managerial context , identify two different and distinct lines of thought on control in the organizational and management literature, which they term 'coercive' and 'caring' respectively. These two lines of research are quite opposed, to the point where they barely engage with each other despite examining, in many cases, the exact same types of management practices. In the *caring* line of research, managerial surveillance and control is understood primarily as legitimated by protecting the majority of employees, by curbing unacceptable behaviour (such as shirking) from a small minority of employees. In contrast, the coercive line of research, has a focus on employees being subjugated by the managerial dominance inherent in systems of surveillance and control. That is, control and surveillance in this respect, is seen as a way to get employees "[...] *work as hard as they can all the time*" (Sewell & Barker, 2006, p. 938), and little else. Sewell and Barker suggest that there is a mutuality to the two lines of research, in the sense that organizational control and surveillance can have varied effects and consequences, both intended and unintended. On one hand, it can foster a counter-reaction by the employees, such as suggested in some of the industrial relations literature, on the other hand, it can help foster co-operation, because of the reliance on procedural fairness and procedural justice are emphasized through transparent rules and regulations.

In the following we will, based discussions above, use a game theory perspective to attempt break down and analyse the relationship between rational and social trust, reciprocity and the employment contract. The purpose of this is to understand how the social motive-based aspects of trust, are related to the underlying social nature of the employment relation. In particular we will differentiate between agency theoretical assumptions of neo-classical

contract theory and the social dynamics of more contemporary approaches, as discussed by Ernst Fehr and his co-authors.

Trust and Reciprocity in the Employment Relation

In the previous section we discussed how trust could be defined, just as we discussed the concept of trust in the employment relation. In particular we also discussed how trust could be in some cases instrumentalized to enable the functionality of modern organic organizations. The aim of this section is to discuss the relation between fairness, trust, and reciprocity through a discussion of two general types of employment contracts, a performance contract and a trust contract. The former kind of contract is accompanied by a tacit assumption on distrust in the sense that it is implied that the employee will not voluntarily choose to exhibit fair behavior. In contrast, the pure trust contract is based on a fairness assumption and as such it is based on the belief that fairness is not a result of steering in either a positive (bonus) or negative (fine) manner. Rather fairness is based on reciprocity. This also implies that egoism cannot be assumed to be the governing relational sentiment, and is thusly separate from the management theories based on overcoming the principal-agent problem.

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[1]. 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 & Gächter, 2000, p. 161).

[1] 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.

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, p. 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, $\pi_i: S_1 X S_2 \rightarrow \mathbb{R}$ (Rabin, 1993, p. 1286). Furthermore, Rabin (1993, p. 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, p. 1286), now, develops a "kindness function" measuring how kind Player 1 is to Player 2 when she chooses a_1 while believing 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 1to Player 2 can now be defined:

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

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, p. 1287) captures in the following function:

$$\tilde{f}_2(c_1, b_2) \equiv \frac{\pi_1(c_1, b_2) - \pi_1^e(c_1)}{\pi_1^h(c_1) - \pi_2^{min}(c_1)} \; if \pi_1^h(c_1) - \pi_2^{min}(c_1) = 0, then, \tilde{f}_2(c_1, b_2) = 0$$

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:

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

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 - implying that she obtains positive utility from hurting the other player.[2] Consider now, a standard Battle of the Sexes:

Figure 1: Rabin's Battle of the Sexes Game (Rabin: 1993: 1285).

		Player 2: Opera	Player 2: 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. Can the different Nash equilibria also therefore also be view as 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

[2] 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.

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.

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

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

$$\tilde{f}_2(O,B) = \frac{\pi_1(O,B) - \pi_1^e(O)}{\pi_1^h(O) - \pi_2^{min}(O)} = \frac{0-2X}{2X-0} = \frac{-2X}{2X} = -1$$

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

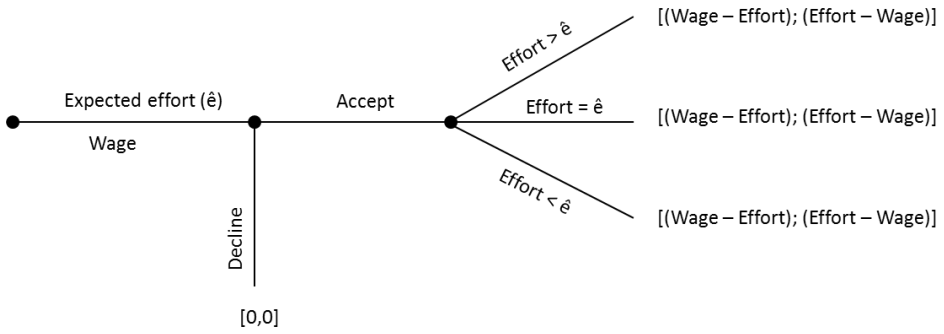
$$U_1(B,B,O) = X - (1-0) = x - 1$$

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, p. 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. 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 & 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 & Fischbacher, 2002, C22, C28). Consider the game below:

Figure 2: Illustration of the game on voluntary cooperation



At the first decision note, 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 principal-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 – ê) 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 & Falk, 2002, p. 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 & Gächter, 2000, p. 171)

This implies two important points: (1) in accordance with Truman Bewley's (1998, p. 475; 2004, p. 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 & Falk, 2002, p. 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 employment 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. This relates to the social motive-based trust, as discussed above, in the sense that it is related to the shared values and background of the employees, as well as the understanding of why someone is doing, what they are doing, outside of the rational and calculative perspective on trust. 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) 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, p. 65; 1991, p. 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, p. 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, p. 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, p. 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, p. 544) then tried to explain this by introducing the notion of gift-giving

inspired by Marcel Maus's (1922/1966) idea on the logic of gift-giving in archaic societies. Maus (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…" (Maus, 1922/1966, p. 9)

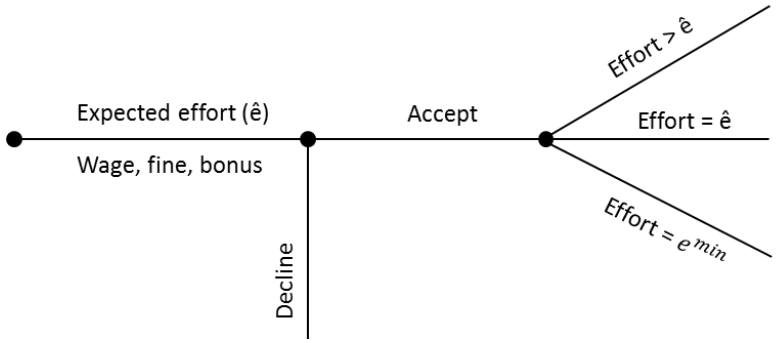
He continues:

> "… 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…" (Maus, 1922/1966, p. 10)

In accordance with this, Akerlof (1982, p. 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 lead Akerlof (1982, p. 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 3: 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, & Königstein, 2008, p. 3). If the employee accepts the offer the following payoffs will emerge:

$$\pi^{Employee} \begin{cases} Accept = w - (7e - 7) \\ Decline = 0 \end{cases}$$

$$\pi^{Employer} \begin{cases} Accept = 35e - w \\ Decline = 0 \end{cases}$$

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, & Königstein, 2008, p. 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, ë*, 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, & Königstein, 2008, p. 4):

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

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

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 = ê.

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

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

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

The game is centered on a similar reasoning as the fine game, above, because the employee will have no incentive to choose e > ê, because then the positive effect of the bonus would be lost which implies that e^{min} is optimal if e different from ê. Hence, the employee will choose e = ê 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 = ê (Gächter, Kessler, & Königstein, 2008, p. 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. 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 (ê-w) the higher effort level supplied (Gächter, Kessler, & Königstein, 2008). 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, & Königstein, 2008, p. 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, & Königstein, 2008, p. 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.

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, & Schmidt, 2007, p. 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, p. 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 an 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, & Schmidt, 2007, p. 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[3] incentive contracts, because the wage offered to the employees was too high or the expected effort

[3] 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.

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, & Schmidt, 2007, p. 131, 134). If these results are viewed in relation to Fehr and Schmidt's (1999) model 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, eventually, to a decrease of the employers' profit (Fehr, Klein, & Schmidt, 2007, p. 146). If the proportion of reciprocal minded employees is 0.4, then:

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

A marginal wage increase yields: 0.4(2/[10+1]) = 0.07 and the employer's gross profit increases by at most: 10x0.07 = 0.7, hence a wage increase is associated with a loss of profit (Fehr, Klein, & Schmidt, 2007, p. 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. 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 payoffs to both parties, and the contracts offered was designed in way which produced low base pay and high bonuses (Fehr, Klein, & Schmidt, 2007). 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, & Schmidt, 2007, p. 141).

The game theory perspective is however limited in various ways. At forefront of this, the models presented above assume that the actors have no previous knowledge of one another. The decisions presented in the models, is therefore taken in a social vacuum, in the sense that the models do not explicitly account for what goes into these decisions. In that sense, the motive-based trust perspective, as discussed in the beginning of the paper, that is, that shared background and values, as well as a tacit understanding of why someone is doing what they are doing, are important to this form of social trust, are not taken into account. Likewise, the distinction between procedural justice and motive-based trust is not easily understood or overcome in this perspective.

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.

Conclusion

The framework presented in this paper draws attention the motives on which the employment relation is based, just as it draws attention to the emotional and social 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. The game theory perspective demonstrates that the trust-based employment contract can function as a form of belief management, where the contract itself signals an intent on part of the manager/employer, in effect showing that the manager/employer implicitly trusts the employee, which is reciprocally returned, in accordance with the labour control perspective, as discussed in the industrial relations literature. The instrumentalization of trust as a management technology then becomes a question of the management of beliefs. That is, the management of beliefs about the intentions of the manager/employer and how that impacts and shapes the relationship with the employee. This showcases, that while there are many and varied benefits to trust-based management (and specifically trust-based contracts), there are also ethical pitfalls in the instrumentalization of trust as a way of shaping behaviour. In short, the instrumentalization of trust can influence the cognitive structure of the relationship, in the sense that it can become manipulation of meaning and perception. This implies, at least to certain degree, that trust can become a matter of enabling social cooperation between the employee and the organization, rather than enabling social interaction in a more general sense, and therefore risks losing its inherent *humanistic* value, as the cement of society or one of the foundations of social interactions in general.

Concluding more generally, the game theory analysis shows us that trust plays a very important role, when it comes to shaping the social foundation of the relation between the manager and the employee, as well as between the employee and the organization (and employer). That is, offering a trust-based contract, which is essentially a guess on the character of the other (i.e. the employee), influences the underlying structure of the game, as presented in the models, which again means that the nature of the relationship changes fundamentally, based on this initial decision. Additionally, this also demonstrates the fickle nature of trust if it is not reciprocated. If the employee offered as trust-based contract, does

not follow the implicit rules or norms, it also changes the nature of future interaction and relationship. Finally, the game theory models also show that reciprocity between employees (that is, the employee-employee relation) is an important factor, underlining the importance of the social environment.

References

Akerlof, G. (1982): Labor Contracts as Partial Gift Exchange. *Quarterly Journal of Economics*, 97(4):543-569.

Aristoteles [Aristotle] (350BC/2007b): *Etikken* [Nicomachean Ethics]. Det lille forlag: Helsingør.

Bewley, T. (1998): Why Not Cut Pay? *European Economic Review* 42:459-490.

Clegg, S.R., Courpasson D. & Phillips, N. (2006): *Power and Organizations*. Sage: London.

Fehr, E. & Falk, A. (2002): Psychological Foundations of Incentives. *European Economic Review* 45:687-724.

Fehr, E. & 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. & Gächter, S. (2000): Fairness and Retaliation: The Economics of Reciprocity. *Journal of Economic Perspectives* 14(3):159-181.

Fehr, E., Gächter, S. & Kirschsteiger, G. (1997): Reciprocity as a contract enforcement device: Experimental Evidence. *Econometrica* 65(4):833-860.

Fehr, E., Klein, A., & Schmidt, K.M. (2007): Fairness and Contract Design. *Econometrica*, 75(1):121-154.

Fehr, E. & Schmidt, K.M. (1999): A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics* 114(3):817-858.

Gächter, S., Kessler, E., & Königstein, M. (2008): Performance Incentives and the Dynamics of Voluntary Cooperation. *Symposium zur Ökonomischen Analyse der Unternehmung.*

Homans, G.C. (1954): The Cash Posters – A Study of a Group of Working Girls. *American Sociological Review* 19(6):724-733.

Katz, D. & Kahn, R. (1966/1978): *The Social Psychology of Organizations.* John Wiley and Sons.

Knights, D. & H. Willmott (1989): 'Power and subjectivity at work: From degradation to subjugation in social relations', *Sociology* 23(4):535-558.

Lewicki, R. J. & Bunker, B. (1995): Trust in relationships: A model of trust development and decline. In B. Bunker & J. Rubin (Eds.). *Conflict, cooperation and justice:* 133-173. Jossey-Bass : San Francisco.

Maus, M. (1922/1966): *The Gift*. Cohen and West LTD.

Organ, D.W. (1997): Organizational Citizenship Behavior: It's Construct Clean-Up Time. *Human Performance* 10(2):85-97.

Rabin, M. (1993): Incorporating Fairness into Game Theory and Economics. *The American Economic Review* 83(5):1281-1302.

Rousseau, D.M., Sitkin, S.B., Burt, R.S., & Camerer, C. (1998): Not so different after all: A cross-discipline view of trust. *Academy of Management Review, 23*(3), 393-404.

Sewell, G. & Barker, J.R. (2006): 'Coercion Versus Care: Using Irony to Make Sense of Organizational Surveillance', *Academy of Management Review* 31(4):934-961.

Simon, H. (1983): *Reason in Human Affairs*. Stanford University Press.

Simon, H. (1991): Organizations and Markets. *Journal of Economic Perspectives* 5(2):25-44.

Tepper, B.J. (2000): Consequences of Abusive Supervision. *Academy of Management Journal,* 43, 178-190.

Townley, B. (1999): 'Nietzsche, Competencies and Ubermensch: Reflections on Human and Inhuman Resource Management', *Organization* 6(2):285-305.

Worshel, P. (1979): Trust and Distrust. In W.G. Austin & S. Worchel (eds.), *The social psychology of intergroup relations* 174-187. Wadsworth: Belmont.

Watson, T.J. (1995): *Sociology Work and Industry.* Routledge: London.