UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Economiche e Aziendali "Marco Fanno"

SCUOLA DI DOTTORATO DI RICERCA IN ECONOMIA E MANAGEMENT
CICLO 25°

FAIRNESS, INTRINSIC MOTIVATIONS AND SOCIAL IDENTITY IN GROUP DECISIONS

Direttore della Scuola: Ch.mo Prof. Giorgio Brunello
Supervisore: Ch.mo Prof. Antonio Nicolò
Co-supervisore: Ch.mo Prof. Rupert Sausgruber

Dottoranda : Diana Gaspari

Contents

Introduzione

In questa tesi ho raccolto tre studi riguardanti l'economia comportamentale in situazioni decisionali di gruppo. Gli esperimenti in laboratorio, con i quali manteniamo il controllo del contesto specifico, ci permettono di isolare i fenomeni oggetto di studio.

Nel primo capitolo, studiamo come la giustizia distributiva influenza la fiducia tra due individui. Una relazione spesso inizia con la divisione di un patrimonio di proprietà comune. La giustizia distributiva di una divisione può avere un'influenza sul successo della relazione. Questo capitolo analizza l'effetto che una procedura di *fair division* ha su un legame di fiducia tra due individui, confrontando due modificazioni del meccanismo Divide & Choose (DC). Nella prima versione del DC, la divisione finale del patrimonio rispecchia le intenzioni sociali del divisore; invece nella seconda versione del DC, a seguito di una divisione equa del patrimonio non è possibile intuire le attitudini egoistiche o sociali del partner. I dati sperimentali evidenziano che, indipendentemente dal meccanismo utilizzato, le coppie che ottengono una divisione equa nella prima fase esibiscono un maggiore livello di fiducia reciproca nella seconda fase. Pertanto, dalla fase di divisione, i soggetti non estraggono e usano razionalmente le informazioni riguardanti le intenzioni sociali del partner. I risultati suggeriscono che una divisione equa crea un atteggiamento positivo, il quale favorisce la fiducia all'interno della coppia, anche nel caso non sia possibile dedurre le predisposizioni sociali dell'altro membro della coppia.

Il secondo capitolo illustra un metodo per misurare la motivazione intrinseca in ambiente lavorativo. In seguito, usando questa misura, approfondiamo come le motivazioni intrinseche si relazionano con gli incentivi estrinsechi. La performance dei dipendenti in un ambiente di lavoro è originata da molteplici fattori. In contrasto con l'economia classica, molti esperimenti economici hanno dimostrato che il denaro non è l'unica motivazione che sta dietro allo sforzo di un lavoratore. In un ambiente lavorativo dove gli individui devono compiere una mansione, è fondamentale per il datore di lavoro capire la motivazione intrinseca verso il compito

affidato al dipendente e l'appagamento personale che quest'ultimo ne può trarre. Abbiamo disegnato un esperimento innovativo che ci permette di misurare la motivazione intrinseca sul lavoro. Nel nostro esperimento, nel quale anche il tempo non lavorativo è salariato, comunichiamo ai partecipanti qual è il tempo ottimale di lavoro. Ciò significa che spieghiamo esplicitamente ai partecipanti come possono massimizzare il loro profitto monetario. La motivazione intrinseca viene definita come la differenza tra il livello di performance osservato e il livello ottimale basato sugli incentivi monetari (estrinsechi) durante lo svolgimento di un compito, in un contesto dove i lavoratori possono essere monitorati. Lo scopo dell'esperimento è di utilizzare la nostra misura di motivazione intrinseca per predire la performance dei lavoratori in un ambiente lavorativo non monitorato. I dati dimostrano che i lavoratori con una bassa produttività hanno una motivazione intrinseca inferiore rispetto ai più produttivi; i lavoratori con un'alta produttività attuano uno sforzo superiore al loro tempo ottimale (in termini monetari) di lavoro.

Infine, il terzo capitolo, analizza l'effetto dell'identità sociale sulla distribuzione di ricchezza. Abbiamo progettato un esperimento, dove sono coinvolti tre individui e due di questi appartengono allo stesso gruppo sociale. Chi ha potere decisionale può scegliere di distribuire equamente la ricchezza a sua disposizione tra i tre individui, oppure può preferire un'allocazione che favorisce se stesso e il soggetto esterno al suo gruppo, creando però uno svantaggio per l'altro membro del gruppo. In questo capitolo studiamo i potenziali determinanti di lealtà verso il proprio gruppo sociale, variando nell'esperimento le identità sociali (basate sul paradigma dei gruppi minimi) e i livelli di status sociale (basati su una misura di abilità cognitive). Inoltre, variamo le opportunità di infliggere punizioni da parte del soggetto che riceve uno svantaggio nella distribuzione della ricchezza. I risultati indicano che persino identità sociali molto deboli hanno un effetto sull'efficienza. Quando una scelta efficiente significa danneggiare il proprio gruppo, gli individui sono propensi a scegliere la distribuzione equa ma inefficiente, perfino supportando un costo personale. Più alto è lo status sociale del gruppo e maggiore diventa il problema dell'inefficienza.

Introduction

This thesis is a collection of three studies concerning behavioral economics in group decision contexts. Laboratory experiments are our main tool to maintain control over the specific settings that we want to analyze and they allow us to isolate the phenomena we are interested in.

In the first chapter, we look at how fairness influences trust between two individuals. A relationship frequently begins with the act of splitting a common endowment. The fairness of this division may influence the success of the relationship. This chapter investigates the effects of a fair division mechanism on an ongoing trusting relationship between two partners, by comparing two different Divide & Choose (DC) procedures. In the first version of the DC mechanism, the division is informative in terms of the pro-social intentions of the divider, whereas in the second version an equal division is consistent both with self-interest and other-regarding concerns. The experimental data find that, irrespective of the mechanism used, couples who reach an equal division in the first phase show higher levels of trust in the second phase. Hence, subjects do not rationally extract and use the information about the intentions of their partner from the division phase. The results suggest that reaching an equal division creates a positive feeling that enhances trust between the couple, even when it is not possible to deduce the intentions of the partner.

In the second chapter, we are able to create a measure of intrinsic motivation in a working environment; and then, using this measure, we investigate how intrinsic motivations interact with extrinsic incentives. The effort of workers in a labor context is motivated by multiple factors. In contrast with the classical economic view, several economic experiments showed that money is not the unique motivation behind their effort. In a working context where subjects have to perform a task, it becomes fundamental for the employer to understand the intrinsic motivation and the taste for the task employees should perform. We develop a novel experiment to provide an empirical measure of intrinsic motiva-

tion at work. In our experiment, with rewarded leisure time, we tell the subjects which one is the optimal working time and consequently how they can maximize their monetary profit. We define intrinsic motivation as the difference between the level of effort actually exerted and the one optimally chosen based on monetary (extrinsic) incentives in a real effort task with complete monitoring of workers. The experiment aims at validating our motivation measure in terms of predictive success in work environments without monitoring of workers. Our data show that low productive workers have lower intrinsic motivation than high productive workers; high productive workers over provide effort working more than their optimal (in monetary terms) working time.

In the third chapter, we analyze the effect of social identity on wealth distribution. We design a three-agent experiment in which two agents belong to the same social group. The decision maker can choose either an allocation that maintains equal payoffs between the three agents, or an allocation that favors himself and the other non-member individual at a cost for the member of his group. We study potential determinants of loyalty by imposing experimental variation in group identity (based on the minimal group paradigm) and status (based on a measure of cognitive ability). Furthermore, we vary punishment opportunities of those who receive a disadvantage from the distribution of wealth. Our results indicate that even very weak group identities have a strong effect on efficiency. When efficiency means hurting the ingroup, subjects tend to choose the inefficient equal distribution even at a personal cost. The higher is the status of the group, the bigger is the inefficiency problem.

Chapter 1

An experimental study of allocation procedures and their effect on trust

Diana Gaspari

I thank the Faculty of Social Sciences (SOWI) of Innsbruck for allowing me to run my experiments in their laboratories. I gratefully acknowledge financial support by the Austrian Science Fund (FWF), under Project no. S 103070-G14.

1.1 Introduction

Two companies interested in forming a joint venture must begin their relationship with a negotiation phase. In this preliminary phase, they negotiate not only how to share the gain of the partnership, but also how to divide their assets if the relationship is terminated. Similarly, in a marriage, spouses often sign a prenuptial contract, which also determines how they split their endowments in the case of divorce.

In this context, it is interesting to study how the outcome of the first negotiation phase affects the parties' incentives to invest in the relationship. The initial negotiation phase, in fact, may not only directly influence the incentives to invest in the relationship to establish the property rights on the future gains, but may also reveal the intentions of the parties and their mutual benevolence. In this work we are interested in analyzing this second aspect, and in particular, whether a fair outcome or a fair behavior is more important to enhance trust in the ongoing relationship.

To this aim, we study experimentally how the division of a common endowment affects the outcome of a subsequent relationship between two subjects. The experiment is composed of two unrelated phases. In the first phase, two subjects divide a common property without being informed that they will play another game in the second phase with the same partner. In the second phase, they play a Trust Game. We design two treatments to distinguish whether it is more effective a fair division, even if it imposed by the rules of the game, or a fair behavior, when agents can choose between picking a favorable outcome or a more equitable division. In the first treatment, we use an allocation mechanism which, in the event of a final fair allocation, the pro-social intentions of the first mover (divider) are clearly revealed. In the second treatment, the fair outcome does not reveal anything about the fair intentions of the divider.

The two allocation procedures provide different information about the motivations behind the other subject's choice. One mechanism is "informative" in the sense that the fair intention for fair division is evident; the other mechanism does not have this property. In this way, we can analyze how the partner's revealed preferences affect the level of investment in a relationship.

Marriages with prenuptial agreements and the allocation of control rights at the beginning of a business partnership are two relevant examples of relationships that are initiated with a division phase. The

literature on assets ownership, starting from Grossman and Hart (1986), mainly studies the effects of the ownership structure on incentives to invest during the relationship. Typically, the ownership structure is decided at the beginning through a negotiation process; in marriages for instance, negotiation takes place while writing the prenuptial agreement. Classical theoretical studies assume that subjects are pure profit maximizers. In this work, we aim at analyzing the role of social preferences in determining the incentives to invest.

Let us consider the marriage example in more detail. Beside the romantic part, there is also the tedious but fundamental issue of the division of tasks and rights between the spouses. Marriage, of course, is an incomplete contract since it is clearly impossible to check and evaluate the effort of the partners in every daily activity. The incompleteness of the marriage contract, together with the increase in divorce rates in the last decades, have caused the spread of prenuptial agreements, especially in the case of second or third marriages (Pilon 2010). The increasing popularity of these contracts have attracted the attentions of some theorists. Rainer (2007) develops a model that determines the equilibrium conditions that assure the efficiency of the prenuptial agreement with respect to the default options of divorce as provided by law. He concludes that equality of divorce payoffs evenly shares the powers within the relationship, since no one can strategically exploit divorce as a threat. Therefore, the incentives for both partners to invest are not undermined by the fear of deception. However, when also considering social interactions, both distributional preferences and intentions influence the behavior of subjects; individuals do not only react to the final outcome of the division, but can also read the motives behind the choices of the other person and respond to them.

Another relevant situation where a relationship begins with the act of splitting an endowment is the business partnership. The companies interested in creating an alliance must negotiate at the onset the control of assets. The result of this first stage is written in a document called a partnership agreement. Only after this initial phase does the business relationship between the two partners begin, and all the characteristics of a working partnership determine the success of the alliance. Joint ventures are spread throughout professional services industries, where joint knowledge can produce significant advantages for both partners. In this particular context, it is clear that trust should be at the inception of the relationship since the individual contribution is quite impossible to monitor. In their empirical study, Bleeke and Ernst (1991) report that joint

ventures with an equal division of control have a higher success rate and a longer life. In contrast, in cases of unequal shares, the majority holder can misuse the greater power and destroy the other partner. They explain that this evidence is based on mutual responsibility. When shares are equal, both parties are responsible for the success of the other; thus, because of this dependency, the likelihood of cheating is decreased. The present study will expand on their ideas by focussing deeply on the allocation process, and on the effect of fairness derived from the first stage within the relationship.

Our work is also related with the literature of signaling and reputation. This wide branch of research introduces reputation concerns to explain costly behaviors that benefit other subjects, such as charitable donations, volunteering, or simply refraining from being selfish. The main idea behind the theoretical models (Benabou and Tirole 2003, 2006; Bernheim 1994; Corneo 1997; Seabright 2004) is that pro-social behaviors are not only driven by distributional preferences, as altruism or inequality aversion, but subjects also value the self-image that they convey towards their social group (Benabou and Tirole 2006). There could be multiple reasons for this; having a reputation of being a pro-social type could be simply a strategy to induce cooperation in subsequent relations, thus increasing the monetary payoff in the long run; or, it could be a means to avoid bad feelings about oneself and enhance one's self-esteem.

Signaling models often refer to subsequent relationships inside the social group of the subject; for example, blood could be donated because of altruism, warm glow or reputation building, even without the need of specifying the next encounter of the donor with another member of the community. Our work deals with exactly these kinds of contexts. Our subjects are in a situation where they are meeting each other in the first phase, but do not know if there will be another encounter in the future. In any case, concerns about one's reputation could have an influence on their choices.

More about this topic can be found in the experimental literature about the spill-over effect between unrelated games. Most of the experimental literature that investigates social preferences mainly focusses on one specific game and on how subjects react inside that particular circumstance. However, when subjects sequentially take unconnected decisions, behaviors could be affected by the previous games even if the situations are unrelated. In the last decade, some papers expanded the research outside a single decisional framing to also consider the spillover effect

among dissimilar games. The most significant research studies the effect of previous decisions on cooperation. For instance, Knez and Camerer (2000) show that coordination in Weak-link Games improves cooperation in a finite horizon Prisoner's Dilemma; Shotter (1999) tests the effect of Minimum and Median Game on a profit sharing contract; and Cason and Gangadharan (2010) find that when the same subjects operate in the same period in a cooperative environment and in a competitive one, cooperation decreases. Other works investigate the spillover effect of learning and rationality. In particular, Mengel and Sciubba (2010) claim that similar games can transfer the learning effect from one to the other. Cherry, Crocker, and Shogren (2003) and Cherry and Shogren (2007) show that rationality, in terms of preference reversals, can be enhanced in a market setting, and that the same individuals are able to transfer it to non-market contexts.

The present work considers a different area of interest. It investigates the spillover effect of information conveyed by fairness on trust between two subjects. A study of similar inspiration, that focuses on the role of information and reputation, is the paper of Albert et al. (2007). In their experiment, they use a first phase where subjects may donate to a charity; the individual donation then contributes to the total amount donated by a predetermined group of subjects. In the second stage, subjects are randomly paired and are only given the aggregate information about the donations of their partner's group in the first phase. At this point, they play a Prisoner's Dilemma or a Trust Game. In both cases, the authors find that there is a spillover effect between the two phases; in particular, they claim that "people are nicer to nicer people". In this frame it is not possible to distinguish if the positive effect of donation comes from the underlying fair intentions of the partner or from the general good feeling generated by fairness. Our work wants to clarify this distinction using two mechanisms that generate fair allocations, but supply different levels of information about pro-social intentions.

The results of our experiment show that "allocation fairness" in this setting is the most important driver of trust independent of information that subjects can acquire from the fair allocation. This suggests that the positive feeling gained by fairness positively influences the level of trust regardless of the information obtained from a fair division. Therefore, even if a fair choice does not mirror a fair strategy, the effect on trust is still positive.

This kind of behavior seems to follow the same rationale of a framing effect. It is well known that the way a situation is described influences

individual's decisions in that moment; for example, Liberman, Samuels, and Ross (2004) show that there is an increase of cooperation in the Prisoner's Dilemma when it is called Community Game rather than Wall Street Game. It could happen that having just experienced fairness creates a generally pleasant feeling that alters behaviors. However in our setting, fairness in the first phase of the experiment should also convey some information and, subsequently, modify beliefs. On the contrary our data do not find a significative role of information about the social preference of the partner in the following Trust Game. An explanation could come from the study of Nagel (1995), which shows that subjects do not possess very deep strategic reasoning, especially in non-repeated games. In the first repetition of Nagel's Guessing Game, the majority of subjects did not exceed iteration step two, where step zero means random play and step one represents the best response to step zero strategies. In light of Nagel (1995)'s results, the inability to use information efficiently could come from a superficial strategic reasoning by the subjects; therefore, to enhance trust in a subsequent relationship, the framing effect of fairness becomes more important than the updating of beliefs.

The chapter is organized as follows: section 1.2 describes the Fair Division mechanisms and the Trust Game used in the experiment, their properties and the theoretical hypotheses; section 1.3 presents the procedural details of the experiment; and section 1.4 shows the analysis of the results.

1.2 The Experiment

The aim of this study is to test how two different Fair Division mechanisms affect trust in a subsequent unrelated interaction between the same couple. In the first phase of the experiment, subjects deal with a Fair Division problem; in the second phase, they play a Trust Game similar to the one presented by Berg, Dickhaut, and McCabe (1995).

The experiment is composed of two treatments which differ in the first phase. In both treatments, subjects are matched in pairs that remain unchanged during the two phases.

1.2.1 Phase one: Allocation Mechanisms

In the first phase of the experiment the members of a couple have to split a jointly owned endowment by means of the so-called Divide & Choose mechanism. This mechanism is composed of two stages: in

Table 1.1: Subjects' values

	subject α	subject β
Good A	$A_\alpha = 9$	$A_\beta = 2$
Good B	$B_\alpha = 2$	$B_\beta = 9$
10 coins (1 coin $c = 1$)	$10c = 10$	$10c = 10$
Total value	$T_\alpha = 21$	$T_\beta = 21$

stage one, the first mover, the Divider, divides the endowment into two portions; in stage two, the second mover, the Chooser, chooses one of the two portions, and the other is assigned to the Divider.

The two treatments of the experiment differ in the method used to assign the roles of Divider and Chooser. In the first treatment, named *RandomDC*, the roles are randomly assigned, and in the second treatment, named *AuctionDC*, they are assigned through an auction. This simple modification fundamentally changes the properties of the mechanism.

In both treatments, the common endowment is composed of two indivisible goods and an amount of coins: good A, good B and 10 coins (single coins are also indivisible). The value of the goods is different for the two subjects, whereas the value of the coins is the same. The goods' evaluation is common information and, therefore, agents play a complete information game. Table 1.1 shows the value of the goods and coins used in the experiment; the subjects' values were preset such that the value of the total endowment was the same for both subjects. Each subject is informed about his own and his partner's value.

Roles assigned randomly - *RandomDC*

In the *RandomDC* treatment, the couple's roles of Divider and Chooser are randomly assigned before the start of the game.

Assuming that the individual utility function comprises only the monetary payoff, the game has a unique subgame perfect Nash equilibrium (SPNE). Proceeding by backward induction in stage two, the Chooser selects the portion that maximizes her[1] personal outcome; in the case of

[1]In this chapter, we use the convention that the first mover is male and the second mover is female.

Table 1.2: All possible proposals of allocation

Portion 1	Portion 2
Good A + Good B + x	10-x
Good A + x	Good B + (10-x)
Good B + x	Good A + (10-x)
x	Good A + Good B + (10-x)

$x \in [0, 10]$=number of coins allocated in portion 1

indifference, she chooses the allocation that maximizes the outcome of the Divider. Anticipating the Chooser's behavior, in stage one, the Divider maximizes his profit by proposing the envy-free[2] allocation that makes the Chooser indifferent between the two portions. In the experiment, the values are set to avoid perfect indifference; therefore, the Divider proposes the division that is closer to the allocation that makes the Chooser indifferent.

Considering our parametrization of the experiment, the outcomes resulting from the SPNE prediction assuming selfish behaviors are:

$$\pi_d = A_\alpha + 8c = 17$$
$$\pi_c = B_\beta + 2c = 11,$$

where π stands for earnings (in experimental currency), and subscripts d and c for Divider and Chooser.

This allocation is ex-post asymmetric, in the sense that the individual outcome depends on the role assigned to the subject.[3] In fact, in the case of complete information, the Divider has an advantage over the Chooser and he can obtain a larger fraction of the value of the entire endowment.

Experimental evidence indicates that subjects are not always interested in the maximization of their own monetary payoff. They also care also about the well-being of others, showing some kinds of social preferences.[4]

[2]The allocation is envy-free if both players do not envy the portion received by the other player.

[3]Due to randomness of roles, the mechanism is ex-ante symmetric in terms of expected payoffs.

[4]See, for example, Bolton and Ockenfels (2000); Cox, Friedman, and Gjerstad (2007); Fehr and Schmidt (1999).

Thus, subjects endowed with other-regarding preferences could systematically deviate from the SPNE prediction based on selfishness. If the Divider is inequality averse, for example, he will choose the allocation that assures the same outcome for both members of the couple; in this case, the outcomes will be $\pi_d = A_\alpha + 5c = 14$ and $\pi_c = B_\beta + 5c = 14$. Beside the equal and the selfish allocations, there are also other plausible no-envy and efficient outcomes that lie in-between the two extremes just mentioned. The full range of possible allocation proposals can be derived from the four cases described in Table 1.2, and the outcome predictions derived from the no-envy solutions when subject α is the Divider (d) as presented in Table 3.3.

In this mechanism, depending on personal social preferences, the Divider could choose one of the efficient and no-envy allocations just presented. Therefore, a selfish Divider will propose the no-envy allocation that maximizes his monetary payoff, and a subject with pro-social intention will choose one of the other solutions according to the shape of his utility function (efficiency requires that good A is assigned to subject α in all these allocations).

The freedom of choice given by the mechanism also conveys some information about the Divider's preferences. By just looking at the proposed portions, the Chooser will infer something about his social preferences. For this reason we will refer to the treatment *RandomDC* as the "informative treatment".

Roles assigned by an auction - *AuctionDC*

The second mechanism solves the ex-post inequality problem of the *RandomDC* when assuming selfishness. In the first stage, instead of a random assignment, subjects have to bid at a first price auction to win the role of Divider. The procedure is similar to the Bidding to be Divider by Crawford (1979).

In this treatment, the roles are determined by an auction rather than by a random device; subject α and subject β submit a bid between 0 and 5 (only integers); whoever wins the auction becomes the Divider. At the end of the game, the amount bid by the winner of the auction (the Divider) will be transferred to the loser (the Chooser). If both subjects bid the same number, a random mechanism determines the winner of the auction.

Table 1.3: Outcome predictions of the allocation phase when subject α is the Divider (d)

RandomDC		
π_d	π_c	
$A_\alpha + 8c = 17$	$B_\beta + 2c = 11$	selfish
$A_\alpha + 7c = 16$	$B_\beta + 3c = 12$	
$A_\alpha + 6c = 15$	$B_\beta + 4c = 13$	
$A_\alpha + 5c = 14$	$B_\beta + 5c = 14$	equal
$A_\alpha + 4c = 13$	$B_\beta + 7c = 16$	
$A_\alpha + 3c = 12$	$B_\beta + 6c = 15$	
$A_\alpha + 2c = 11$	$B_\beta + 8c = 17$	altruistic
AuctionDC		
π_d	π_c	
$A_\alpha + 8c - b_\alpha = 17 - 3 = 14$	$B_\beta + 2c + b_\alpha = 11 + 3 = 14$	selfish
$A_\alpha + 8c - b_\alpha = 17 - 3 = 14$	$B_\beta + 2c + b_\alpha = 11 + 3 = 14$	equal

The *RandomDC* version of the Divide & Choose shows that in the unique SPNE with selfish players, the Divider earns 17 and the Chooser 11. Thus, the Divider has an advantage of 6 over the Chooser; this amount also represents the maximum rational bid in the auction stage of any SPNE of the mechanism when subjects are selfish.

Starting from the assumption of maximization of the monetary payoff, in this second game there is a unique SPNE. In stage two, the Chooser takes the portion that maximizes her monetary payoff; in the case of indifference, we assume that she chooses the allocation that maximizes the outcome of the Divider. In stage one, the Divider, knowing the behavior of the Chooser in stage two, maximizes his profit by proposing the envy-free allocation that makes the Chooser indifferent between the two portions. In the auction stage, both subjects bid an amount that corresponds to half of the Divider's advantage of the allocation stage and a random mechanism assigns the roles.

The outcomes resulting from the selfish SPNE prediction are presented in Table 3.3. It is clear from the analysis of the equilibrium outcomes that this second mechanism guarantees ex-post equality even when both subjects are selfish.

However, as previously mentioned, the assumption that all subjects want to maximize their monetary payoff is questionable in this kind of game. A Divider strongly inequality averse, for instance, will equally split

the endowment. Therefore, the optimal bid depends on the preferences of the player.

It is easy in equilibrium to check that inequality averse subjects will bid 3 points in the auction stage. Then, if he becomes the Divider, he will propose the allocation that, after the transfer of the bid, guarantees equality, which is exactly the same proposal of a selfish subject, as we can see in the lower part of Table 3.3. If, instead, he becomes the Chooser, he is sure that, even when facing a selfish Divider, the final allocation will be equal.

Contrary to the *RandomDC*, this mechanism guarantees ex-post equality even in the case of selfish subjects, but on the other hand, the Chooser cannot make any inference about the Divider's preferences. The Divider can no longer propose different divisions that mirror his selfish or pro-social intentions. The mechanism *AuctionDC* does not permit the acquisition of information about the Divider's social preferences. For this reason, the *AuctionDC* treatment will be named "uninformative treatment".

Assuming that the population is composed of subjects with different social preferences, we can draw some hypotheses about the degree of equality that the two mechanisms, *RandomDC* and *AuctionDC*, can guarantee. First of all, we define the equality degree as the difference between the payoffs of the two partners; we find the highest degree of equality when the difference is zero and the two partners earn exactly the same amount of points. The larger is the difference, and the lower is the degree of equality.

The mechanism *AuctionDC* provides incentives for both selfish subjects and those with pro-social preferences to choose an allocation that equally splits the value of the endowment. On the contrary, in the mechanism *RandomDC*, Dividers with pro-social intentions choose the equal allocation, whereas those who are selfish propose allocations with a lower degree of equality. Therefore, considering the totality of the subjects, the two mechanisms should generate different degrees of equality when played by a sample that comes from a population with different types of subjects.

In our first hypothesis, we look at the difference in the payoff of each Divider/Chooser couple in the two treatments. Let $\overline{D}_t$ be the average difference of the monetary payoff of the couples in treatment t for $t = \{r = random, a = auction\}$.

H1: The average difference $\overline{D}$ is lower in treatment *AuctionDC* than in treatment *RandomDC*:

$$\overline{D}_a \leq \overline{D}_r$$

Hypothesis H1 wants to verify the degree of equality that the mechanisms can reach when applied to a random sample of the population. The difference between the outcome of the Divider and the outcome of the Chooser, named D, is the measure of equality of the mechanisms. In this setting, $D = 0$ represents the highest degree of equality; in fact, when the difference is equal to zero perfect equality is guaranteed. The higher D is the bigger the inequality is between partners, thus raising unfairness. Hypothesis H1 claims that the average D is smaller in the *AuctionDC* mechanism since in this game all different types of subjects head toward the same equal outcome.

1.2.2 Phase two: Trust Game

In the second phase of the experiment, subjects play a Trust Game (Berg, Dickhaut, and McCabe 1995). This subsequent phase of the experiment is unrelated to the first one, i.e. subjects during the first phase only know that there will be a second phase, but they do know what it will be about.

The Trust Game is a sequential game that involves two subjects, the Trustor (first mover) and the Trustee (second mover). Both subjects are endowed with an amount M of money, but only the Trustor can decide how much to invest (transfer) $T \in [0, M]$ by giving it to the Trustee. The amount transferred is then tripled $T * 3$ and the Trustee now decides how much to give back to the Trustor $X \in [0, 3T + M]$.

The profit functions of the Trustor (r) and of the Trustee (e) are respectively:

$$\pi_r = M - T + X$$
$$\pi_e = M + 3T - X$$

The SPNE of this game, assuming that subjects maximize their monetary payoff, predicts $T = 0$ in stage one; in fact, anticipating a selfish behavior of the Trustee $X = 0$, the Trustor does not invest anything. Dropping the assumption of selfish behaviors, we could observe some positive transfer $T > 0$, that measures our level of trust of the Trustor.

The subjects in this second phase of the experiment are again assigned the same partner from phase one. However, in this phase subjects make decisions in both roles (strategy method), i.e. they act both as Trustor and as Trustee.

A subject can face two different situations depending on his/her and the partner's role in the allocation phase (the first phase of the experiment). In the event that the subject was the Divider in phase one, and now he assumes the role of Trustor, he has no information about the social preferences of the Chooser/Trustee. Nevertheless, if in phase one the Divider has been "fair" to the Chooser offering more than the selfish SPNE prediction, he might anticipate positive reciprocity of the Trustee (Chooser), which induces him to invest more in the Trust Game. If, instead, the subject was the Chooser in phase one and now she assumes the role of Trustor, she could observe the Divider's proposal and, depending on the treatment, infer something about the social preferences of the partner. If in phase one the Divider has been "fair" to the Chooser, the Trustor (Chooser) now can decide to invest more, since the information on the Trustee's pro-social attitude helps to form positive beliefs of trustworthiness.

The existing literature on spillover effects between unrelated games suggests that we should observe some differences in the level of trust between the two treatments. The different fairness and informative properties of the allocation mechanisms of phase one permit the identification of which elements in the tradeoff between equality and information about the intentions are more important in creating a trusty environment inside the couple.

In the following hypothesis, named H2, the terms *fair*, *self* and *altr* indicate strategies that lead to outcomes. For example, in the *RandomDC* mechanism, a selfish strategy leads to an unequal outcome, and in the *AuctionDC*, a selfish strategy generates the same equal outcome as pro-social strategies. We categorize the Divider's no-envy proposals[5] in the three groups "selfish", "fair" and "altruistic" following these criteria: in "fair" are the proposals that almost equally split the endowment (the difference between the monetary values is less or equal than two); "selfish" are the proposals that induce a difference in the payoff greater than two in favor of the Divider and, when the Chooser has an advance

[5]We only consider no-envy proposals because in equilibrium we should never observe an envy proposal. Indeed, in the experiment, 96% of the Divider's proposals (62 out of 70) satisfy this property.

of more than two points on the Divider, the proposal is "altruistic".

Let us define trust T as the amount invested in stage one of the Trust Game by the Trustor, and the average level of trust of treatment t as $\overline{T}_j$ for $t = \{r = random, a = auction\}$.

H2: Conditional on the Divider's proposal being "fair", the average amount invested by the Trustor is higher in the *RandomDC* (r) treatment than in the *AuctionDC* (a) treatment:

$$\overline{T}_{fair,r} > \overline{T}_{fair,a}$$

Conditional on the Divider's proposal being "selfish", the average amount invested by the Trustor is higher in the *AuctionDC* (a) treatment than in the *RandomDC* (r) treatment:

$$\overline{T}_{self,a} > \overline{T}_{self,r}$$

Conditional on the Divider's proposal being "altruistic", the average amount invested by the Trustor is equal in the two treatments:

$$\overline{T}_{altr,a} = \overline{T}_{altr,r}$$

The level of trust depends on the beliefs that the Trustor forms looking at the outcomes of the Divide & Choose. However, these outcomes are produced by the choice of a strategy that is not directly observable. The two treatments differ in the degree of information about social preferences that the subjects acquire while observing the proposed and the accepted allocation in phase one.

If a fair outcome is generated in both mechanisms, pro-social intentions are identifiable only in *RandomDC*. The outcomes coming from *AuctionDC* do not provide information about the social preference type of the subjects, thus beliefs on trustworthiness are not influenced from the *AuctionDC* allocation mechanism: $\overline{T}_{fair,a} = \overline{T}_{self,a}$. If the outcome of the Divide & Choose is determined by a selfish strategy, the lowest level of trust is generated by *RandomDC* since the selfish intentions are clearly revealed. In the case of altruistic proposals, both treatments will reveal information and, thus, have the same effect on trust.

The alternative hypothesis predicts no spillover effects between the two phases of the experiment. In this case, the two decisional processes

are completely independent, and fairness deriving from the allocation phase has no influence on trust in the second phase.

The three cases presented in hypothesis H2 assume that the Chooser always behaves efficiently, this means she chooses the portion containing the more valuable good for herself. However, in the case of a Divider's unfair proposal, we could also expect some sort of punishment by the Chooser. In situations where punishment is verified, the information conveyed from phase one to phase two is not any clearer. Suppose that the Chooser observes a selfish proposal and punishes the Divider; then, in the role of Trustor she can believe either that the Divider/Trustee becomes even more selfish to recover the losses suffered in phase one, or that the punishment was successful in explaining the social norm of fairness and, therefore, it helps cooperation. Hence, we decide to exclude from the analyzed data the few cases where punishment occurred (8 couples out of 70), since the design of the experiment is not suitable to understand the effect of punishment on trust and it is not the aim of our work.

1.3 Experiment's Details

Every subject is matched randomly with another of the same session and the couple does not change in the two phases of the experiment. Maintaining the same couples, we can show the direct effect of the allocation mechanism on a subsequent relationship with the same partner.

The experiment consists of two treatments with a between subjects design. The treatments are imposed in the first phase, i.e. each group of subjects plays one of the two allocation procedures followed by a common phase where they are involved in the Trust Game.

The results of both phases are added up to form the final payment of subjects at the exchange rate of 1 point = 0.50 euro.

The experiment was run at the experimental laboratory of the University of Innsbruck (Austria) between November 2011 and January 2012 with 140 subjects recruited via ORSEE (Greiner 2004). Subjects were students from different faculties with no exclusions. The experiment was programmed and conducted with the software z-Tree (Fischbacher 2007).

1.3.1 Phase one: Allocation Mechanisms

In the first phase of the experiment, each couple of subjects has to play one of two versions of the allocation mechanism Divide & Choose.

The two treatments differ from each other in this phase.

Subjects are randomly paired and the two members of each couple are randomly chosen to be subject α or subject β. The game is played one-shot.

Treatment *RandomDC* assigns the roles randomly. The experimenter goes to each subject with a bag containing numbers from 1 to 20. Each subject draws one number and types it into the computer under the supervision of the experimenter to avoid falsehoods. The member of the couple that picks up the higher number becomes the Divider. In treatment *AuctionDC*, before the division stage, both members of the couple bid an amount of points between 0 and 5. This range has been decided to avoid excessive losses in the experiment and it contains the equilibrium bid of 3. Whoever wins the auction becomes the Divider, and at the end of this first phase, she/he has to transfer the amount bid to the Chooser.

After the assignment of roles, the Divider forms two portions out of the common endowment and sends the decision to the Chooser. On the screen the subjects can clearly see the personal evaluation of the endowment for both members. The Chooser now clicks on the portion that she/he wants to have. At the end the result and a summary of the choices are shown to both subjects. This feedback gives information about the strategies the subjects used, and from this observation, they may or may not change their beliefs about the social preferences of their partner.

1.3.2 Phase two: Trust Game

In the second phase of the experiment subjects play a Trust Game. The game is the same in the two treatments. After the end of the Divide & Choose phase, the instructions for the Trust Game are handed in to subjects in order to exclude any reputation or signaling strategy in the first phase.[6]

At this time, each subject receives her/his show-up fee of 10 points (5 euros); these personal endowments are used to play the Trust Game. The nature of the game permits each subject to preserve entirely the show-up fee in case of distrust, or to increase it in case of trust and a trustworthy response. Thus, we feel confident using their show-up fee in this phase because any loss is a personal choice and they are perfectly

[6] At the beginning, the experimenter tells the subjects that the experiment is composed of two phases and the instructions are explained at the beginning of each phase.

aware of the risk. An equal initial endowment, moreover, excludes any choice driven from pre-game inequality concerns.

The Trust Game is presented to subjects using the strategy method. The task is very simple to understand and subjects can quickly and easily make choices to use the strategy method confidently. After some control questions, every subject decides how much to transfer in the role of Trustor.[7] At this point, we also ask for the elicitation of beliefs, i.e. in the same screen there is a table that the Trustor fills in with her/his belief about the back transfer of the Trustee contingent on every option of the Trustor. The specific question is: "If you send to the Second Mover ... points, how much do YOU expect the Second Mover will send back to you from her/his amount at her/his disposal (show-up fee + your transfer tripled)?" Our methodology permits us to collect more information, thus obtaining a different function of beliefs for every subject.

In the second screen, subjects make their decision in the role of Trustee. They see a table with 11 entries, one for each possible level of investment of the Trustor (from 0 to 10), and they have to choose how many points to send back to the Trustor from every possible amount at their disposal (show-up fee + First Mover's investment tripled).

At the end, a random mechanism decides the effective role of the two subjects and their choices determine the final earnings of this phase. The results of the two phases are added to establish the final payment.

1.3.3 Questionnaire

While waiting for payments, subjects are asked to fill in a questionnaire. Besides the classic socio-demographic profile, we added the questions of the German Socio-Economic Panel Study (SOEP) survey (Naef and Schupp 2009) in order to detect betrayal aversion and risk attitudes, which are the other two main drivers of trust together with beliefs. In this way, we have the complete picture about which element of our analysis generates trust in the second phase.

1.4 Results

This section presents the results of the experiment. In the first part, data from the two versions of the Divide & Choose are analyzed, and second, we describe the interaction between the allocation mechanisms

[7] In the instructions and during the experiment the Trustor is called "First Mover" and the Trustee "Second Mover".

Table 1.4: Details of subjects and earnings

Treatment	subjects	independent obs.	average earnings (euro) (std. dev. in brackets)
RandomDC	60	30	14.07 (3.89)
AuctionDC	80	40	13.99 (3.90)
Experiment	140	70	14.02 (3.88)

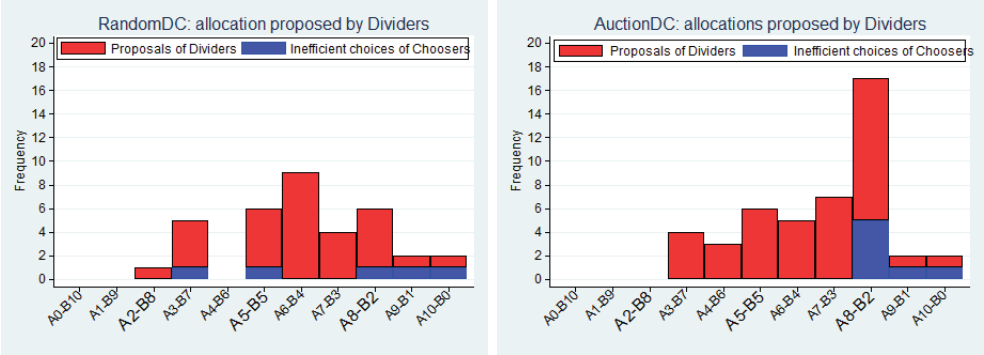

Figure 1.1: Allocation phase: distribution of Dividers' proposals: treatment *RandomDC* on the left, treatment *AuctionDC* on the right. The x-axis orders the plausible divisions of the Dividers holding fixed the division of the two goods and varying the partition of coins. We assume that *good A* and *good B* are in two different portions and *A* is the efficient good, which means the good with the highest value for the Divider.

and the Trust Game. In the experiment, 140 subjects participated, 60 in treatment *RandomDC* and 80 in treatment *AuctionDC*. The average earning was 14.02 euro (Table 3.4).

1.4.1 Results of Phase One: Allocation Mechanisms

Figure 1.1 shows the distribution of the Dividers' proposals of the two treatments in the allocation phase of our experiment. On the horizontal axis, we list the possible divisions of the endowment.[8] In Figure 1.1 we

[8]The whole range of possible divisions also includes the case where the two goods are in the same portion; for example, good A and good B in one portion and the 10 coins in the other portion. Only one Divider opted for this kind of allocation and we decided for simplicity of representation to exclude it from the graph, but not from the data analysis.

assume that the Divider proposes an efficient division, in the sense that he intends to have the portion with the more valuable good.

The blue bars are the number of inefficient choices of Choosers after the proposal of that particular allocation. Specifically, here inefficiency means that after the proposal of two portions by the Divider, the Chooser takes the portion with the good that has the lowest personal evaluation, thus minimizing the total size of the cake.

From Figure 1.1, we note first of all the high willingness to maximize the total value of the endowment of the subjects: only 12 couples (blue bars) out of 70 failed to reach the efficient division.

In the middle point of the horizontal axis, the allocation A5-B5 equalizes the payoffs of the two members of the same couple; on the left side proposals give a higher payoff to the Chooser and on the right side to the Divider. Proposals from A2-A8 to A8-B2 are no-envy. Outside this interval at the extremes of the graph the proposals are envy, and the Chooser has a monetary incentive to choose the inefficient allocation, thus increasing the likelihood of inefficient choices.

Looking at the two distributions, we note that the greater part of the division proposed by the Dividers lies between the equal allocation A5-B5 and the selfish one of A8-B2; in particular, 76.67% of the proposals in *RandomDC* and 76.92% in *AuctionDC* fall into this range. The altruistic divisions represent only minor fractions, 16.66% in *RandomDC* and 17.95% in *AuctionDC*. The two distributions of the Divider's proposals do not statistically differ between treatments (Mann-Whitney test $p < 0.3855$ and Kolmogorov-Smirnov test $p < 0.611$).

The two mechanisms are characterized by two different methods of role assignment; the *AuctionDC* exploits the transfer of the bid to compensate for the theoretical advantage of the Divider when we assume that subjects are selfish. Dividers in this second treatment should behave in the division stage anticipating the final transfer of the bid, and the two stages cannot be considered as independent. Thus, the similarity of the two distributions of proposals of the division stage seems inconsistent with our theoretical predictions. If the two subjects sample come from the same populations the distribution on the right graph in Figure 1.1 should be shifted towards the more selfish proposals given that in the final stage the points bid by the Divider will be transferred to the Chooser.

Figure 1.2 represents again the distribution of the Dividers' proposals in treatment *AuctionDC*; in this graph, however, we adjust the proposals considering the transfer of the bid from the Divider to the Chooser. For

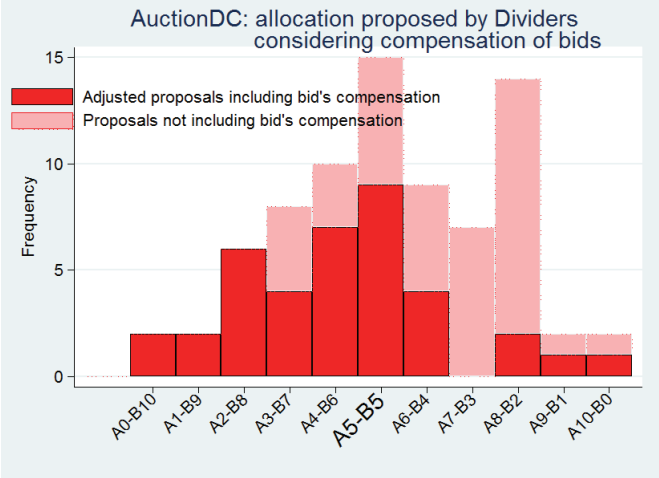

Figure 1.2: Allocation phase: distribution of Dividers' proposals; treatment *AuctionDC*. The x-axis orders the plausible divisions of the Dividers holding fixed the division of the two goods and varying the partition of coins and considering the compensation deriving from the payment of the bid to the Chooser by the Divider. We assume that *good A* and *good B* are in two different portions and *A* is the efficient good, which means the good with the highest value for the Divider.

example, if the proposal is A7-B3 and the payment due to the Chooser is 2, the allocation presented in Figure 1.2 will be included in the category A5-B5 (A7-2; B3+2). The lighter colored bars in the background report the proposal's distribution without the compensation on bids, in other words the same distribution of the right graph of Figure 1.1.

When considering the payment of bids in the proposals, we notice a relevant shift of the distribution towards the left, and it is also statistically different from the distribution of proposals in treatment *RandomDC* (Mann-Whitney test $p < 0.0002$ and Kolmogorov-Smirnov test $p < 0.001$). The shift means that in *AuctionDC*, 55.88% of the proposals are in favor of the Chooser (allocation on the left of the equality proposal A5-B5) against the 16.67% in *RandomDC*. Considering previous studies on distributional preferences (Forsythe et al. 1994) and the relevant difference with respect to the *RandomDC*, it seems unrealistic that in the *AuctionDC* more than half of the Dividers have altruistic preferences. The main difference between treatments is the strategic complexity, i.e. subjects have to iterate in *AuctionDC* one step more than in *RandomDC*. Our explanation for the shift of the distribution lies on this characteristic of the design. We see in Figure 1.1 that behaviors in the division stage are not different and they mirror the social preferences of the Dividers.

It seems that some subjects do not consider the auction stage and the division stage as a unique game, and they are not able to anticipate the payment of the bid while dividing the endowment in two portions. Therefore, when the Dividers have to pay the bid to the Choosers at the end of *AuctionDC*, they end up in a disadvantaged situation most times. This explains why the majority of Dividers in *AuctionDC* are victims of a "winner's curse" earning at the end less than their partners.

In summation, the two mechanisms do not differ much in terms of equity contrary to hypothesis H1. The main difference between the two mechanisms concerns unequal allocations; in *RandomDC* inequality is in favor of the Divider as the theory predicts, whereas in *AuctionDC* the Chooser has the advantage in the case of an uneven division.

We shift the attention now to the main aim of our study, which is the effect of the degree of equity and the information about intentions on trust between the same subjects. However, before proceeding into the analysis, we have to define at what point we determine an allocation is equal. Following the reasoning of H2 and given the wide range of possible divisions, we decided to classify as "equal" not only perfect equality, A5-B5, but also the neighboring proposals A4-B6 and A6-B4. All the other allocations are defined as "unequal".[9]

1.4.2 Result of Phase Two: Trust Game

The second part of the experiment wants to find the spillover effect of the division mechanisms on trust. The main result is that trust is influenced by equality deriving from the allocation phase, but not by the information acquired about the social preferences of the partner, i.e. "intentional fairness" does not affect trust.

We define trust as the amount of money transferred by the Trustor (first mover) in the Trust Game. Figure 1.3 compares the distribution of trusty decisions splitting the observations between treatments and with

[9]Hypothesis H2 considers three categories of proposals: fair, selfish and altruistic. In the data analysis, we decided instead to group the observations in only two categories: equal and unequal. There are two main reasons for this: i) the number of observations for the altruistic category is too limited to be included in the analysis (6 Dividers out of 30); ii) the few allocations deriving from an altruistic decision of the Divider in the informative treatment *RandomDC* has a negative effect on trust. One explanation could be that inequality aversion in this setting is considered as a better predictor of subsequent trusting behaviors than altruism. For these two motivations we decided to pool the data of "selfish" and "altruistic". However, results are robust even when the "altruistic" category is excluded.

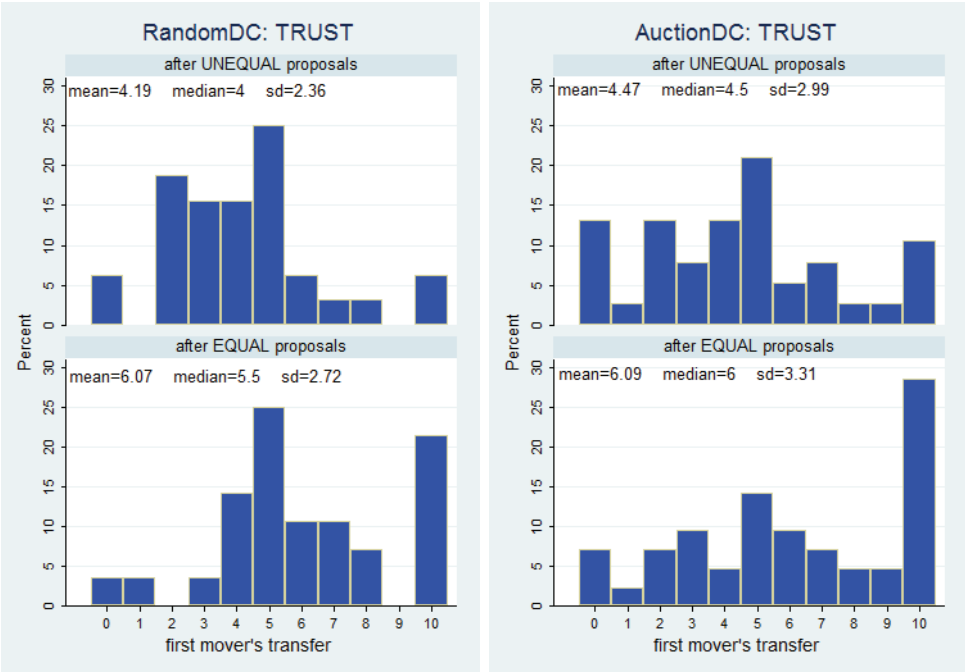

Figure 1.3: Trustor's transfer in the Trust Game sorted by treatment and by proposal's fairness.

Table 1.5: Average trust (std. dev. in brackets)

Treatment	After equal proposals	After unequal proposals
RandomDC	6.07 (2.72)	4.19 (2.36)
AuctionDC	6.09 (3.31)	4.47 (2.99)

respect to the classification of equality deriving from the allocation phase. On the horizontal axis, we find the 11 levels of Trustor's investments from points 0 to 10.

The graphical analysis excludes the couples who experienced the inefficient choice of the Chooser when the Divider proposes a no-envy division; this means, for instance, that the Chooser takes the portion with good A when the proposal is between A2-B8 and A8-B2. In this range, the portion containing good B always has a greater monetary value for the Chooser; if the Chooser in this situation takes the portion containing good A, she is clearly bearing a cost to punish the partner. It would be interesting to see the effect of punishment on trust, but only eight couples (three in *RandomDC* and five in *AuctionDC*) ended up in this category and the limited number would not give us any insight. The effect of punishment will be included in the regression analysis.

We start the analysis of the results by considering the two treatments separately and we look first at the treatment *RandomDC* (the left column of Figure 1.3). In this treatment, trust increases on average from 4.19 to 6.07 when the Divider proposes an equal division in the allocation phase (Mann-Whitney test $p < 0.0038$ and Kolmogorov-Smirnov test $p < 0.069$). The difference in the informative treatment *RandomDC* proves a clear positive effect of equality coming from an unrelated preceding decision task on trust.

In the informative treatment *RandomDC*, we cannot distinguish if the positive effect on trust of the Divider's proposals in the "equality" category is due to the objective "allocation fairness" or to the acquired information about the intentions of the partner (intentional fairness). To distinguish the two kinds of fairness we now consider all subjects that experienced an equal proposal and we compare the two treatments (the lower row of Figure 1.3). Keeping in mind hypothesis H2, the level of trust should be the lowest in "*RandomDC* after an unequal proposal" since the intentionality of selfishness is clearly revealed; trust in "*AuctionDC* after an equal division" should be in the middle position because

there is no updating of beliefs and "*RandomDC* after an equal division" should produce the highest level of trust. The analysis of data on the contrary does not support our hypothesis H2. The two distributions on the lower part of Figure 1.3 do not statistically differ (Mann-Whitney test $p < 0.9274$ and Kolmogorov-Smirnov test $p < 0.744$), and we can even observe a peak (29%) in the efficient but very risky solution $Trust = 10$ in the category "*AuctionDC* after an equal division". It seems that subjects create a positive judgment about their partner only looking at the superficial fairness of proposals and not reading the intentions behind them.

We have seen that in both treatments having an equal allocation in the first phase has a positive effect on trust. This evidence in treatment *AuctionDC* indicates that the positive effect of equality on trust does not come from the understanding of pro-social intentions, but rather from the objective "allocation fairness" independently of what the mechanism reveals about the preferences of the partner.

Table 1.6 reports the estimation results. Model (1) is an OLS regression with *Trust* as the independent variable that is defined as the amount transferred by the Trustor to the Trustee. Following the reasoning of Fehr (2009), in this first regression we check the three main determinants of trust: beliefs, risk and betrayal aversion. The variable *Beliefs* is composed of the amount that each Trustor guesses that the Trustee will transfer back conditional on his transfer; i.e. if the Trustor decides to transfer five points to the Trustee, we will consider the particular Trustor's belief about the back transfer of the Trustee only concerning the initial transfer of five points. *Risk Attitude* and *Betrayal Aversion* are measures derived from the final questionnaire. In this first model we control for gender (*Female*) and for the role of the allocation phase (*Divider*). In the setting of our experiment, only *Beliefs* have a positive and significant effect on *Trust*; in particular, increasing the belief of back transfer by one point, we observe an increment in Trust of 0.415 points. Risk and betrayal aversion do not have an effect.

Referring once again to the work of Fehr (2009), the treatment variation should influence Trust through beliefs rather than directly. This is exactly the aim of models (2) and (3) where the independent variable is *Beliefs*. Model (2) tests the treatment effect indicated by the dummy variable *AuctionDC* and the effect of having experienced an equal proposal of division in the first phase of the experiment with the dummy variable *Equal Division*. Model (3) also includes the interaction term

	(1) Trust	(2) Beliefs robust s.e.	(3) Beliefs robust s.e.	(4) Trust	(5) Trust
Beliefs	0.415***			0.410***	0.417***
	(0.000)			(0.000)	(0.000)
Risk	-0.00664			-0.0177	-0.0208
	(0.892)			(0.714)	(0.670)
Betrayal	0.127			0.119	0.126
Aversion	(0.153)			(0.173)	(0.153)
Female	-0.410	-1.496	-1.479	-0.465*	-0.457*
	(0.102)	(0.145)	(0.150)	(0.057)	(0.063)
Divider	-0.185	-1.621	-1.620	-0.200	-0.201
	(0.448)	(0.119)	(0.120)	(0.398)	(0.400)
AuctionDC		1.572	0.928	-0.452*	-0.0483
		(0.113)	(0.483)	(0.062)	(0.917)
Equal Division		3.118***	2.369*	0.416*	0.788**
		(0.003)	(0.077)	(0.098)	(0.043)
Punishment				-0.906**	-0.546
				(0.025)	(0.468)
AuctionDC			1.311		-0.643
x Equal Division			(0.511)		(0.222)
AuctionDC					-0.292
x Punishment					(0.749)
AuctionDC					-0.00868
x Beliefs					(0.842)
Cons	1.369***	8.162***	8.503***	1.665***	1.396***
	(0.002)	(0.000)	(0.000)	(0.000)	(0.007)
N	140	140	140	140	140

p-values in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.6: OLS-Regression. In models (1), (4) and (5), the first mover's transfer is used as a measure of Trust. *Beliefs* represents the belief associated with the amount transferred by each subject. *AuctionDC* is the treatment dummy. *Equal Division* is a dummy variable that identifies the equal proposals in the allocation phase. *Punishment* is a dummy variable that indicates members of the couples where the Chooser punished the Divider. The variables *Risk Attitudes* and *Betrayal Aversion* come from the questionnaires. The variables *AuctionDC x Equal Division*, *AuctionDC x Punishment* and *AuctionDC x Beliefs* are the interaction terms. In models (2) and (3), *Beliefs* is the dependent variable and robust standard errors are used.

between *Equal Division* and the treatment dummy *AuctionDC*. In both regressions, we can see a clear positive effect of equality derived from the previous phase of the experiment on the Trustor's beliefs, but there is not any treatment effect. The two treatments provide different information conditional on the equality of the division in the allocation phase. However, from the regression analysis it seems that subjects do not use this information to modify their beliefs about their partner's decision. As we have seen in the nonparametric analysis in Figure 1.3, "allocation fairness" is the unique characteristic of the division that subjects consider while forming beliefs about their partner's social preferences.

In models (4) and (5), the independent variable is again *Trust* and beyond *Beliefs*, *Risk Attitude* and *Betrayal Aversion*, we test for the the direct effect of treatments (*AuctionDC*), equality of proposals (*Equal Division*) and *Punishment*. The dummy variable *Punishment* refers to those couples where the Chooser decided to renounce to some money in order to leave to the Divider the portion with the good that had the lowest personal evaluation for the Divider, i.e. the Chooser chose the inefficient allocation. Comparing models (1) and (4), we can see that the positive effect of *Beliefs* also remains controlling for the direct effect of an equal division on trust. The coefficient of *Equal Division* decreases considerably between the two models and also becomes less significant. Model (4) confirms the fact that the variable *Equal Division* affects trust mainly through beliefs. The treatment dummy *AuctionDC* shows a weakly significant negative effect that disappears when controlling for interaction terms in model (5). The negative effect indicates that trust in the case of unequal proposals is lower in treatment *AuctionDC*. The possible explanation could be that an unequal proposal in treatment *AuctionDC* creates difficulties in understanding the intentions of the Divider, since it is a strategy that brings, in our case, an advantage to the Chooser, thus easily attributable to a mistake of the Divider. In any case, the result is not robust when including interaction terms. Couples that experienced *Punishment* are inclined, not surprisingly, to reduce the level of trust toward the partner, but again the effect disappears when controlling for interaction terms in model (5).

Now let us turn our attention to the beliefs functions of the subjects. Analyzing them in detail, we observe an interesting behavior of their standard error variation. In the informative treatment *RandomDC*, the standard deviation of beliefs on trustworthiness is similar between the equal and unequal Divider's proposals of the allocation phase (variance-

comparison tests for every beliefs' distributions of each possible first mover's transfer). This equivalence is coherent with the transparency of the mechanism about the social preferences of the partner. In the case of either an unequal or equal proposal, the Chooser infers the intentions of the Divider and forms beliefs accordingly, thus uncertainty remains constant.

AuctionDC instead records a higher variance of beliefs in the case of an unfair proposal by the Divider. An unequal allocation in *AuctionDC* creates some confusion in the Chooser's mind since unequal proposals in this treatment come from the difficulty of the Divider to recover the initial bid in the division stage; this kind of "error" seems to leave subjects in a fog. However, the variance in case of a equal proposal in *AuctionDC* is similar to the one in *RandomDC*. This means that equal proposals decrease uncertainty until they reach the same level of *RandomDC*.

Once again, the positive framing of fairness has an effect independently of the information that fairness provides. In this case, the bias takes place in the beliefs' variance. Fairness reduces uncertainty even in the uninformative treatment where there are no sufficient reasons to explain this fact.

The standard deviation's difference between fairness and unfairness becomes evident when we consider the equality of the final allocation, and not the proposal of the Divider. In this circumstance, the variance of beliefs is lower in the case of fairness for both treatments. The final allocation is the last screen that the subjects see before playing the Trust Game; therefore, the good feeling given by fairness is more important in the trusting behavior than the information that can be culled from the Divider's proposal. It seems as though they are unable to use the information to update their beliefs, but they stop thinking when they know the final allocation without analyzing how that particular outcome was reached.

1.5 Conclusion

The allocation mechanisms, studied in the bargaining and the Fair Division literature, are built to satisfy particular properties of fairness. However, if the allocation stage takes place at the beginning of the relationship, we should ask ourselves whether the specific properties of the mechanisms also have influence on the success factors of the relationship. We considered, in particular, the cases of marriages and business partnerships, where it is necessary to build a trusting environment to enhance

specific investments. For this reason this chapter tests the spillover effect of two Fair Division mechanisms on trust between the same subject couples.

The experiment compares the influence on trust of a Fair Division mechanism that is theoretically unfair when assuming selfishness but provides information about the social preferences of the other subjects, with a mechanism that guarantees fairness leading subjects toward the same fair allocation but gives up the possibility to read the pro-social attitudes of others.

The main result of our study is that fairness matters independently of the information deriving from the equal allocation. The perception of a sense of fairness in the air acts similar to a positive framing effect. This generally optimist feeling increases the likelihood of trusting behaviors toward other people.

Returning to the paper by Bleeke and Ernst (1991) cited at the beginning of this study, we can say that fairness in the allocation stage contributes to the success of the partnership not only because of mutual responsibility as claimed by the authors, but also because the positive feelings conveyed by equality increases trust between the parties even in unconnected sequential relationships. Therefore, in the case of a relationship that begins with the division of some property rights, the planner of the allocation mechanism should place a lot of attention on the fairness of the final allocation to create a constructive atmosphere where the connection can begin in a solid way.

Chapter 2

A measure of intrinsic motivation in the workplace

Diana Gaspari
Rupert Sausgruber - WU Vienna, Institute for Public Sector Economics

We thank the Vienna Centre of Experimental Economics(VCEE), University of Vienna, and the Faculty of Social Sciences (SOWI) of Innsbruck for allowing us to run our experiments in their laboratories. We gratefully acknowledge financial support by the Austrian Science Fund (FWF), under Project no. S 103070-G14.

2.1 Introduction

Workers, while performing a task, are motivated both by extrinsic motivations, as salary, and by intrinsic motivations, for example their taste for the task. Extrinsic motivations can be influenced by the employer via different pay-schemes or benefits. Intrinsic motivations instead are more heterogeneous; therefore difficult to capture and even more complicated to steer.

When we talk about intrinsic motivation in this work, we refer to the broad definition from organizational literature.[1] Intrinsic motivation is the "the desire to engage in an activity because one enjoys, or is interested in, the activity" (Sheldon et al. 2003).

Our study aims at finding a measure of intrinsic motivation. We implement a real effort laboratory experiment where subjects have to work for 20 minutes. The task becomes more and more difficult to perform during the 20 minutes, and subjects have the possibility to stop working and switch to paid leisure. When the task is too difficult and it takes too much time, it is more profitable for the subject to switch to leisure. In this way we are able to estimate an optimum switching point for each subject that depends on individual ability and productivity. The distance from the optimum switching point and observed behaviors of participants is our measure of intrinsic motivation. Results show that low productive subjects have a lower intrinsic motivation than high productive subjects and more productive ones oversupply effort showing a high intrinsic motivation.

Intrinsic motivation has been considered deeply in psychology literature[2] already in the 70s; subsequently also economists started to investigate how it interacts with performance and effort and in particular the focused on the crowding out phenomenon. Frey (1992) was the first to use the psychologists' distinction between extrinsic and intrinsic motivation to explain why incentives may have counter intuitive effects. Later, Frey and Jegen (2001) developed the Motivations Crowding Theory to explain formally how intrinsic motivation can decrease due to external incentives and Benabou and Tirole (2003) defined also the specific setups in which crowding out may occur.

Experimental literature[3] also focused on the effect on performance

[1] See Rosso, Dekas, and Wrzesniewski (2010) for a discussion about different kinds of motivations on the workplace.

[2] See Ryan and Deci (2000) and Leonard, Beauvais, and Scholl (1995) for an overview of psychology literature.

[3] See Gneezy, Meier, and Rey-Biel (2011) for an overview of experimental literature.

due to external incentives. When performance did not vary accordingly to maximization of monetary income, authors referred to a modification of intrinsic motivation. One well-known example is the paper of Gneezy and Rustichini (2000), where they showed that monetary incentives in voluntary contexts work in the desired direction only above a certain amount, small incentives undermine intrinsic motivation. The work of Segal (2012) is closely related to our study. She showed the presence of intrinsic motivation using unincentivized simple tests. She was able to connect the performance of the tests to the success in the labor market. Using really easy tests, cognitive abilities were not prominent and therefore the main predictor for future achievement were intrinsic motivation.

Even if intrinsic and extrinsic motivations were treated in several fields of studies, as far as we know, a quantitative measure of intrinsic motivation is still missing. With our work we are able to quantify intrinsic motivation and to untie it from cognitive capabilities. This work will be the base for further research on the effect of extrinsic incentives on intrinsic motivation. In the second part of the chapter, we present an example using a team pay-scheme as extrinsic incentive.

In Section 3.2 we explain the design of the experiment. We present some theoretical predictions and the estimation model in Section 3.3. The details of the experiment are explained in Section 3.4 and subsequently we report the results in Section 3.5. In Section 2.6 we bring an example of how extrinsic incentives affect intrinsic motivation of workers.

2.2 The Experiment

The aim of this study is to find a measure of intrinsic motivation in the workplace. At the scope we design a real effort laboratory experiment.

The real effort task lasts for 20 minutes. During this time, subjects can generate income by solving cross sums at their own speed. These cross sums are presented to participants on a computer screen, where they have to fill in the correct answers into blanks. Each of these screens contains three cross sums. Once all cross sums on a screen are correctly solved, the participant can proceed to the next screen. If a mistake is made, the program will highlight the incorrect answer and rectification is mandatory before carrying on is possible. Cross sums become more difficult over time. More precisely, exercises consist of two digits at the beginning of the experiment, and an additional digit is added after every 15^{th} correctly solved task (or every fifth screen).

During the 20 minutes, besides the cross sums task, subjects also have

the opportunity to opt for the so-called SWITCH-task at any time.[4] Yet, once subjects opt out of the cross sums task they can not revise this decision for the remainder of the experiment.[5]

For each correctly solved cross sum participants receive a piece rate of 0.7 points, while in leisure mode, participants earn 1 point per 15 seconds spent in this modus.

The key issue of our design is the increasing difficulty of cross sums, and consequently the increasing opportunity cost of calculation represented by the payment during leisure mode. This design results in an optimal switching point, which depends on individual productivity, and moreover we are able to avoid corner solutions. In general, a profit-maximizing subject should switch once it takes her more than 31.5 seconds to correctly solve one screen of cross sums. This is clearly communicated to participants via the instructions.

The payoff function, given our parametrization, is for subject i:

$$\pi_i = 0.7 * cross\,sums_i + \frac{leisure\,time_i}{15}$$

2.3 Theoretical Predictions

Assuming that participants in this experiment are rational profit-maximizers, we are able to formulate the following profit function:

$$\pi_i = w\ f(e_i, \eta_i) - c(e_i) \tag{2.1}$$

Where π_i denotes the individual profit and $c(e_i)$ represents the opportunity costs of participation in the real effort task. The first term on the right side of equation (2.1), $f(e_i, \eta_i)$, represents the subject's production, which is a function of individual time allocated to the real effort task e_i and individual productivity η_i, which is assumed to be constant over time.[6] The wage rate, w, is also fixed at 0.7 points per correctly solved cross sum. We further assume that marginal productivity with respect to work effort is positive but diminishing, i.e. $f'(e_i, \eta_i) = \frac{\partial f}{\partial e_i} > 0$ and

[4] In the instructions the real effort task is referred to as the "cross sums task" while the leisure mode is named "SWITCH-task".

[5] To prevent accidental switching, individuals have to check a confirmation box in addition to pressing the SWITCH-button.

[6] This assumption is not unrealistic given the results presented below. We also found no reaction on the intensive margin in another set of experiments previously conducted.

$f"(e_i, \eta_i) = \frac{\partial^2 f}{\partial e_i^2} < 0$. Maximizing (2.1) with respect to work effort e_i we derive the optimality condition:

$$w\ f'(e_i, \eta_i) = c'(e_i) \tag{2.2}$$

Consequently, subjects should provide work effort until their marginal costs, which are represented by their opportunity costs, equal their marginal return from solving cross sums.

An important characteristic of our design is that the optimal switching time varies according to the individual ability and it is unique for each individual. Based on our experimental design, which consists of increasing difficulty and opportunity costs, we assume effort to be a nonlinear function of cross sums solved of the form:

$$T_i = \alpha_i + \beta_{1i} Q_i + \beta_{2i} Q_i^2 \tag{2.3}$$

Where T_i denotes the time spent on the real effort task, Q_i represents the number of cross sums solved by subject i, β_{1i} and β_{2i} are productivity parameters and α_i is a constant. In general we can rewrite equation (2.1) as:

$$\pi_i = w\ Q_i + l\ (E - T_i) \tag{2.4}$$

where Q_i remains the number of solved cross sums by individual i and l represents the wage rate in the SWITCH-task, which is fixed at 1 point every 15 seconds. E is the endowment of time (20 minutes) and T_i is again individual i's time spent on the cross sums task. Considering the nonlinear relationship of T_i and Q_i our profit-maximization leads to:

$$\pi_i = w\ Q_i + l\ [E - (\alpha_i + \beta_{1i} Q_i + \beta_{2i} Q_i^2)] \tag{2.5}$$

Derivation of (2.5) with respect to Q_i results in the FOC:

$$Q_i^* = \frac{1}{2\beta_{2i}} (\frac{w}{l} - \beta_{1i}) \tag{2.6}$$

Q_i^* is our benchmark to measure intrinsic motivation. Given that we tell clearly to subjects that to maximize their payoffs they should switch to leisure when it takes more than 31.5 seconds to solve one screen, deviations from the optimum switching point should reflect the intrinsic

motivation of subjects. Therefore intrinsic motivation M_i is the difference between the observed number of cross sums Q_i and the estimated optimum number of cross sums Q_i^*:

$$M_i = Q_i - Q_i^* \tag{2.7}$$

2.4 Experiment's Details

In total 196 students, recruited via ORSEE (Greiner 2004), participated in this experiment. The experimental sessions were conducted at the computer laboratory of the Faculty of Social Sciences (SOWI) of Innsbruck and at the Vienna Centre of Experimental Economics (VCEE) between March 2010 and December 2013 and lasted for approximately 90 minutes each. The software employed for programming and conduct of the experiment is z-Tree (Fischbacher 2007).

Before the start of each session, individuals were randomly assigned to a seat, and subject-IDs were randomized to assure absolute anonymity of participants. Subsequently, instructions containing the information were handed out and participants had sufficient time to study these instructions thoroughly. The identities were not revealed at any time, and any form of communication was strictly prohibited.

Prior to the real-effort phase, subjects had to answer some questions to make sure that they fully understood the design of the experiment. Following these questions, a trial phase was conducted to familiarize subjects with the game environment and the cross sums task. The trial round lasted for 240 seconds and no payment was made for the achievements during this time. Following the training round, the 20 minute real-effort phase started as described above.

At the end of the experiment we pay subjects at the exchange rate of 15 point = 1 €.

2.5 Results

In this section we present the results of the experiment. As mentioned above, 196 subjects participated in our experimental sessions. Average earnings were around €14.50 and no show up fee was paid.

The number of solved cross sums was on average 99.30, it ranged from 15 to 159 with a median value of 102 (Figure 2.1).

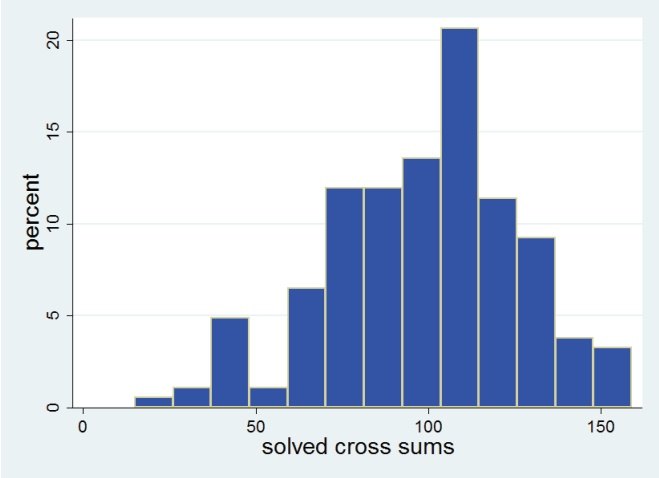

Figure 2.1: Distribution of solved cross sums.

Individual data permit us to estimate the individual optimum number of cross sums and the optimum working time to maximize the monetary profit, using the model described in Section 3.3. Each participant has a personal optimal point and we will use this reference as a benchmark to analyze the behaviors of our subject pool.

In Figure 2.2 we see how much subjects deviate from their estimated optimal points. Subjects are ordered on the horizontal axis according to their productivity, i.e. the number of cross sums solved before switching to leisure, and on the vertical axis we plot the difference between the estimated optimum number of solved cross sums and their effort (observed number of solved cross sums). Subjects close to the zero line maximize their monetary profit; subjects in the lower part of the graph solved too few cross sums and the ones in the upper part over provide effort.

Data show a clear positive relationship between productivity and distance from the optimum point (correlation=0.6659, p-value=0.0000). The low productive subjects work less than the optimum point; when productivity increases effort approaches the optimum point, and the highest productive subjects over provide effort.

> Result 1: The low productive subjects under provide effort and the high productive subjects over provide effort with respect to their individual optimum point.

Before claiming that the distance to optimality reflects intrinsic motivation, we have to exclude that the effect is due to cognitive capabilities.

Figure 2.2: Intrinsic motivation. On the horizontal axis observations are sorted by number of cross sums, the vertical line represents the median point 102. On the vertical axis it is represented the difference between the estimated individual optimum number of cross sums and the observed one.

It could be possible that low productive subjects under provide effort because they do not understand well incentives. Therefore we asked to 56 of the participants to perform an IQ test before the cross sums tasks.[7]

Table 2.1 reports the estimation results. We use the number of cross sums solved as dependent variable. First of all, we notice in the estimation that our measure of intrinsic motivation has a positive highly significative effect on the performance of subjects; this confirms our Result 1. The IQ test score has no impact on performance and it is not correlated with motivation (corr=-0.0662 p-value=0.6341).

> Result 2: Cognitive capabilities do not explain the distance to the optimum number of cross sums that maximizes the monetary payoff.

The design of our experiment is able to capture the intrinsic motivation of participants. Subjects know clearly their optimum switching point,[8] nevertheless we observe lots of deviations, that means that they

[7]The treatment with the IQ test last for 20 minutes more than the other treatments. The payment function did not vary, but at the end of the experiment we increased the payments of participants of 5 Euros each.

[8]We explain explicitly in the instructions how to maximize their monetary payoff.

	OLS	
	Cross sums	
Intrinsic motivation	0.663***	(0.000)
Female	-6.902*	(0.197)
IQ	0.747	(0.068)
Working hours per week	0.029	(0.878)
Cons	76.39***	(0.000)
N	54	

p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.1: OLS-Regression. The dependent variable *Cross sums* is the number of cross sums solved. The variable *Intrinsic motivation* is the difference between the individual optimum and the observed number of cross sums. *Female* is a dummy variable for the gender and the variable *Working hours per week* indicates the how many hours participants usually work during the week.

are maximizing a utility function that is not composed only by the monetary payoff, but there is also a component of intrinsic motivation that is not explained by cognitive capabilities.

2.6 Team incentive schemes

In the second part of our work, we start to study how extrinsic incentives interact with intrinsic motivations. The number of possible extrinsic incentives is vast and we cannot consider all of them in this chapter. However we decided to start here with a team incentive scheme to give the flavor of the interaction effect.

We decided to use this kind of extrinsic incentive, since in the last decades grouping workers in teams has become a common organizational instrument (Bandiera, Barankay, and Rasul 2012; Hamilton, Nickerson, and Owan 2003), however empirical and experimental literatures have not yet paid much attention on the theme. Teams in working environments are powerful instruments for firms that want to gain efficiency merging different individual skills of workers, but at the same time teams introduce the temptation for free riding.

80 of our participants had to work for other 20 minutes on the same cross sums task. In this second phase they were matched with another participant. The second phase is equivalent to the first phase but instead of earning 0.7 points per correctly solved cross sum, individuals also receive 0.35 points for each cross sum their partner solves correctly. Yet, these positive external effects only exist as long as subjects spend time in the cross sums task. Once a participant opts for the SWITCH-mode, her counterpart does not receive any further externality payments. In general, the existence of positive externalities does have an important implication as it changes the design into a kind of Public Good Game. In such a case the rational switching point from a pro-social and efficiency-enhancing perspective will be shifted upward, from 31.5 seconds to 47.25 seconds per screen. This is communicated to participants via the instructions.[9] It is important to highlight that the optimal switching time for a selfish individual, however, remains unchanged with respect to phase 1.

The payoff function of the second phase, given our parametrization, is for subject i:

$$\pi_i = 0.7 * cross\,sums_i + 0.35 * cross\,sums_j + \frac{leisure\,time_i}{15}$$

where j is the other member of the group.

In order to exclude the learning effect as explanation of the change in behaviors between the two phases, 60 of our participants had to perform a second real-effort phase identical to the first one, without team incentives. We refer to this control treatment as BASE. Treatment with the team incentive is called TEAM.

In the first phase, where participants worked individually, we saw that low productive subjects have also a lower intrinsic motivation, and the opposite for the high productive ones. In this setting, team incentive schemes modify the intrinsic motivation of our participants.

The graphs of Figure 2.3 maintain the same individual productivity classification on the horizontal axis (number of cross sums in phase 1). On the vertical line we have our measure of motivation in phase 1 in the upper part and in phase 2 on the lower part. The two graphs on the left refer to treatment BASE and the two on the right to TEAM.

[9] At the beginning of the experiment, participants knew that there will be more phases but they did not know the details of the next phases. Instructions for the second real-effort phase were handed out only after the conclusion of the first phase.

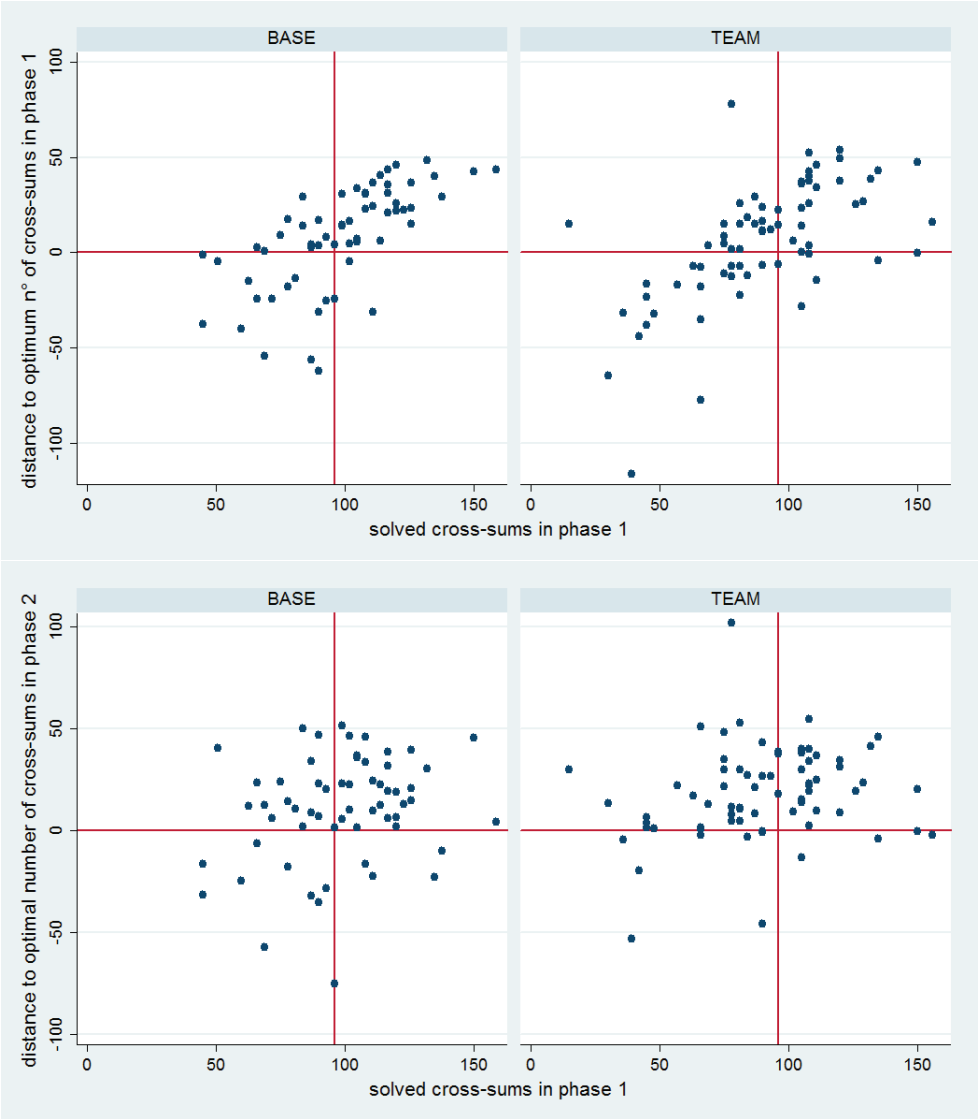

Figure 2.3: Effect of an extrinsic incentive on intrinsic motivation. On the horizontal axis observations are sorted by number of cross sums solved in phase 1, the vertical line represents the median point 96 (value for the data of the two treatments BASE and TEAM). On the vertical axis it is represented the difference between the estimated individual optimum number of cross sums and the observed one. In the upper part of the figure the difference is relative to the observed number of cross sums in phase 1 and in the lower part to the observed number of cross sums in phase 2.

As we already saw in Result 1, in phase 1 (the first two graphs in the upper part), data show a clear positive relationship between productivity and distance from the optimum point in both treatments (Chow test, p=0.4169[10]).

The two graphs in the lower part of the Figure 2.3 maintain the same productivity classification on the horizontal axis and look at the behavior of subjects in the second phase. Here, we notice clearly that the high productive subjects do not change behaviors between the two phases, whereas who was a low productive subject in phase 1 increases the effort in phase 2 even above the optimal individual point. This effect is more pronounced in treatment TEAM (Chow test, p=0.0853).

> Result 3: In phase 2, the low productive subjects increase the effort even beyond their individually optimal point; the high productive subjects instead do not change their level of effort. The increase of effort of the low productive subjects is significative larger in treatment TEAM than in treatment BASE.

The introduction of an extrinsic incentive has an effect not only on performance, but also on the intrinsic motivation especially of the low productive subjects. An important implication of this result is that, when we have heterogeneous workers, extrinsic motivations provided by team incentives do not have the same effect on all subjects. In our context, they have an effect only on the lower productive subjects. High productive subjects have such a high intrinsic motivation that additional incentives have not the power to increase it further. We should take in account that, for the subjects that already find a motivation untied to money, changing the pay scheme may be not really effective.

2.7 Conclusion

Intrinsic motivation on the workplace plays a fundamental role on productivity. Given that intrinsic motivation is a really subjective driver of behaviors, it is extremely difficult to identify.

[10]The Chow test controls the equality of coefficients of two linear regression with two distinct data sets. In our case the two data sets are observations from the two treatments, BASE and TEAM. The equations used for the test is $Q_i - Q_i^{self} = \alpha + \beta_1\ Q_i + \beta_2\ treatment + \beta_3\ Q_i X treatment$ and then we test whether β_2 and β_3 are equal to zero.

With our work we are able to measure quantitatively the intrinsic motivation of workers. Using a control lab experiment with a real effort design, we keep high control but at the same time we obtain results that are applicable to real world situations.

Data indicate that most of subjects do not maximize their monetary payoff. Our design permits us to estimate the individual optimum point and we measure intrinsic motivation comparing the observed behaviors with this point. We observe a relation between productivity and motivation. Low productive subjects solve less cross sums not only for lower personal ability in calculations, but also for a lack of intrinsic motivation. On the contrary, high productive subjects over provide effort even beyond their optimal point.

This work could be considered the base to study the effect of job design on the intrinsic motivation of workers. In the last part of the chapter we bring an example of how extrinsic incentives may modify the intrinsic motivation. Introducing a team pay scheme, we see that the incentive has an effect on intrinsic motivation of only the low productive subjects. High productive subjects, that show already an high intrinsic motivation, do not change their behaviors due to monetary incentives.

Chapter 3

Acting White: an experimental approach

Diana Gaspari
Rupert Sausgruber - WU Vienna, Institute for Public Sector Economics

We thank the Vienna Centre of Experimental Economics(VCEE), University of Vienna, for allowing us to run our experiments in their laboratories. We gratefully acknowledge financial support by the Austrian Science Fund (FWF), under Project no. S 103070-G14.

3.1 Introduction

In migration and job flows, self-selection and signaling tools are important mechanisms to improve efficiency in allocation of workers. Through signaling devices, as education for example (Spence 1973), employers are able to assign high skilled jobs to more capable applicants. High educated also self-select in migrations and sort themselves in countries with significant wage difference between high and low-skilled workers (Grogger and Hanson 2011).

Social identity, however, can shrink the potential efficiency of these mechanisms. Social groups usually develop an identity that creates a moral obligation of loyalty towards the group. When an individual tries to move socially upwards, for example investing in education, the social group could perceive the choice as a betrayal to their identity.

In this chapter we test, with a laboratory experiment, whether attachment to social identity could generate inefficiency in a society. In our experiment the "society" is formed by three subjects and one participant has the power to choose between two distributions of wealth. The decision maker is linked to another participant via social identity, the third one represents the outgroup. One distribution assigns the same wealth to all three participants; the second one is unequal since both the payoffs of the decision maker and of the outgroup subject are larger than the ingroup subject's payoff; therefore the ingroup mate is damaged, and moreover the total wealth (efficiency) is higher.

The data acquired from the lab experiment show that when subjects are linked by a social identity, they prefer to choose the equal but inefficient distribution not to hurt the ingroup mate. Therefore in our setting the attachment to a social identity generates inefficiency in the society.

With this design we want to represent the situation where a capable individual can choose between move socially upwards, betraying his social identity (he generate a damage for the ingroup, since it loses a skilled individual), and be faithful to the ingroup decreasing the overall efficiency.

One prominent application of the context just described is the phenomenon called "Acting White". The expression "*Acting White* describes a set of social interactions in which some minorities incur costs for investing in behaviors characteristic of whites (e.g., raising their hand in class, making good grades, or having an interest in ballet)" (Fryer and Torelli 2010). The costs derive from the perceived betrayal by the minority and the consequently exclusion from the social group. In the United

States, it has been widely observed that black students perform worse at school with respect to their white mates; the "Acting White" theory is one explanation that attracted several researchers in the last years. The contexts, where the "Acting White" phenomenon could be relevant, are really widespread, however the empirical studies present some shortcomings and there are no control experiment as far as we know.

In Section 3.2 we explain the design of the experiment. Theoretical predictions are presented in Section 3.3. Subsequently we report the details of the experiment in Section 3.4 and results are finally discussed in Section 3.5.

Literature review

The inefficiency problem, that rises from avoiding to move socially upwards due to social identity, has been studied mainly relative to the "Acting White" phenomenon. The concept "Acting White" started to be mentioned in the economic literature from the middle 80's. The seminal paper of Fordham and Ogbu (1986) is the first work to use the term; they argue that the phenomenon is caused by a scarce self-esteem of black student derived from decades of racial discrimination.

Only after some years, a few empirical studies tried to understand better the motivations that lie behind this behavior. Cook and Ludwig (1997) and Ainsworth-Darnell and Downey (1998) are the first to expand their analysis to a nationally representative sample. The weakness of these data base is that identity is measured by means of a self reported indication of popularity.

Fryer (2006) and Fryer and Torelli (2010) were able to introduce an objective measure of social connection. They asked to students to mention their closest friends. Then looking at the frequency of names it was possible to understand who is more or less popular inside the community in point. Contrary to previous studies they found clear evidence that blacks and whites link differently social status to academic achievements.

These studies shed light on inefficiency generated by social groups. However, empirical data are not able to disentangle the pure effect of social identity from the potential economic benefits of the network. With our experiment we are able to exclude any other possible advantage derived from the social group and we isolate the effect generated by the attachment to social identity. As far as we know, this is the first work that applies a laboratory experiment to this context.

In the experimental literature our work can be connected to the research on social identity related to status and to redistribution. Already in the 70s social psychologists understood that individuals link themselves deeply to social groups. Tajfel and Turner (1979) developed the Social Identity Theory to find a motivation for discrimination between groups. This theory claims that social identity starts to take shape categorizing people into different groups according to diverse criteria as race, religion, gender and so on. In the second step, the subject includes himself in those categories and thirdly his group is compared to the others. At this point the subject assigns an affective meaning to his group and he cares not only about himself but also about the wellbeing of the whole group.

In the last 20 years social psychologists performed several experimental studies to observe how social identity influences decisions both inside the group and toward external groups. The method mainly used is the minimal group paradigm (Tajfel and Turner 1979); with which researchers prime natural identities or create groups artificially. The main insight from this literature is that social identity is a fundamental decisional driver of individual and leads to ingroup favoritism and outgroup discrimination.

More recently economists started to pay attention to how social identity affects economic decisions. The seminal paper of Akerlof and Kranton (2000) develops the concept of identity-based utility. The degree of attachment to a particular social group influences decisions in many economic areas, as for example in the labor market or politic voting. In their model, one subject forms beliefs about the social norm of his group, and his utility decreases when behaviors are not in compliance with these norms. Currarini, Jackson, and Pin (2009), in their theoretical model, moved the attentions to "homophily". In their opinion, the utility of being part of a group derives from the mere feeling of friendships. Social identity has been also experimentally related to strategic settings, allocation tasks, endogenous group-formation and the tradeoff between more identities. Results are still mixed and depend on the way identity is formed. In some experiments identity is primed through some questionnaires, in others it is created by means of some tasks in the laboratory, or alternatively experimenter went on the field to find and use existing identities.[1]

The experiments more related to this chapter study the effect of social identity on redistribution. Klor and Shayo (2010) used natural identities

[1]For more details about the experimental literature see Chen and Li (2009).

(field of study) to study the voting behavior over redistribution. They created two groups and assigned them two distributions of wealth; one group was on average poorer than the other one. Moreover inside each group some subjects were richer or poorer with respect to the average wealth of their group. Each subject then had to choose between two tax rates. Their design created a conflict for rich (poor) subjects in the poor (rich) group; for example a rich subject in the poor group who wanted to maximize his profit should have voted for a low tax rate, but if he cared about the well being of the group he should have voted for a high tax rate. They showed that identity makes subjects in conflict less selfish favoring thus the total payoff of the group. In this work however, subjects had to choose how to redistribute a certain wealth, but there wasn't any loss of efficiency whatever the choice is. Our experiment instead introduces the conflict between efficiency and social identity.

A natural extension to these studies is to use directly wealth as identity. The Social Identity Theory (Tajfel and Turner 1979) claims that strength of identification with a social group depends on the status of that group. The members of a group feel more attached to it if the status is high. Shayo (2009) explicitly included status in his theoretical model. The two main component for identification are status and similarity; he claimed that "given these two components, an individual is said to identify with group J if (1) he or she cares about the status of group J and (2) he or she wants to resemble the members of group J". As a consequence, if we consider for example different levels of income as social groups and wealth as a signal for status, rich subjects should feel more identified since their status is higher.

Several experimental papers proved that this claim is often verified. Roccas (2003) created status first through preference over paintings where groups were defined "analytical" or "holistic"; then by a speed task, where groups were denominated "time speeder" or "time delayers". It was clearly explained to subjects that analytical and time speeder groups are generally more successful in work contexts and also in personal relationships. They found out that subjects in the high status groups both perceive the higher status and identify deeper with the group. Guimond, Dif, and Aupy (2002) proved the same effect creating groups of "nondisabled" students (regular students) and "disabled" students (students with important learning problems at the end of the primary school). In the paper of Klor and Shayo (2010) already cited above, it was showed also that subjects in rich groups have a higher probability to identify with it. In general every individual has more identities

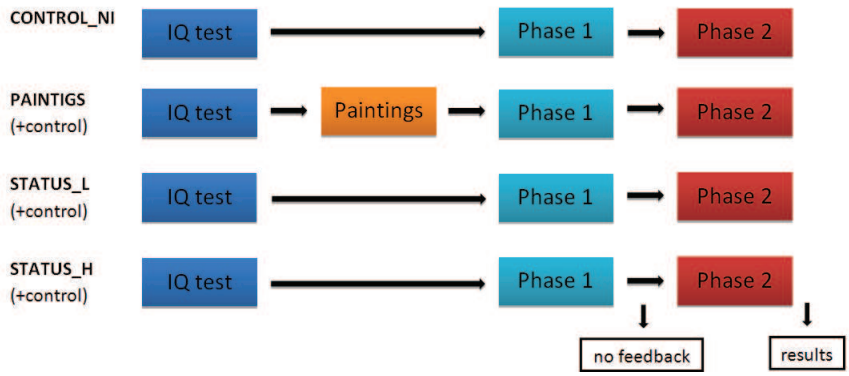

Figure 3.1: Experimental design

at the same time and some identities could prevail depending on the context. Charness (2012) created two identities in a Public Good Game, one with a team-building task[2] and one simply giving different endowments. He found out that subjects with higher endowments tend to aggregate together, reducing the link to the group made through the team-building task. Beside the experimental evidence, Shayo (2009) brought also some empirical evidence to support his theoretical work. Both at the individual and at the national level,[3] he found that individuals with lower income tend not to identify with the low income group but rather being nationalistic.

3.2 The Experiment

The experiment is composed by four main treatments where we vary the strength of the social identity, and by three control treatments (Figure 3.1). Each treatment has two parts; in the first part identity is created and in the second part subjects are clustered in groups of three and they have to decide about the distribution of wealth among the members of the own group. We vary the treatments in the first part modifying identities.

The identity part of treatment ***Paintings*** consists of two phases. The first is an IQ test.[4] This IQ test will determine the identity of sub-

[2]Each team was supplied with a set of letters and subjects together had to form words.

[3]International Social Survey Program (ISSP) 1995 - National Identity surveys, World Values Survey (WVS), and the Luxembourg Income Study (LIS)

[4]The Raven Advanced Progressive Matrices Test.

jects in treatment *Status*, therefore to maintain control we decided to implement the test in all treatments. It does not provide any earnings to participants. In the second phase, we create the identity of the groups using the method based on the minimal group paradigm of Tajfel et al. (1971). Subjects visualize two paintings, one made by Kandinsky and one by Klee, in five consecutive screens.[5] In each screen they have to indicate which painting they like the most, without specifying who the author is. According to the preferences over the two artists, subjects are split in two categories: participants who prefer Kandinsky and participants who prefer Klee. After the identity part, subjects are matched in groups of three; the groups remain unvaried for the whole experiment. Groups are formed by two subjects that like the same artist, participant 1 (P1) and participant 2 (P2), and by one subject that prefers the other artist, participant 3 (P3); for example one group can be composed by P1=Kandinsky, P2=Kandinsky and P3=Klee. In the second part of the experiment, subjects have to decide how to distribute wealth inside their group. This second part is composed by two decisional phases. In the first one, they have two choices at disposal: distribution LEFT gives 20 points to all three subjects and distribution RIGHT gives 16 points to P1, 26 points to P2 and 26 points to P3. As we can see from the distributions, our design creates a conflict for P2, who must choose between an efficient[6] distribution that decreases the payoff of his "identity partner" P1 and increases his own payoff and the one of P3, and an inefficient distribution that is costly for himself and for P3 but does not harm P1. Subjects only know the compositions of their group (Kandinsky-Kandinsky-Klee or Klee-Klee-Kandinsky) and the artist they prefer; therefore P1 and P2 do not know their roles. We ask P1 and P2 to take their decision in the role of P2 and we tell them that they will discover their role at the end of the experiment. P3 obviously recognizes his role and we ask him expectations about the decision of P2 in his group.[7] At the end of this first decisional phase they do not receive any feedback. The second decisional phase is identical to the first one in the decision of distributions, but we

[5]The 5 couples of paintings are: *Mountain Formation*, 1924, by Klee and *Aglow*, 1928, by Kandinsky; *Warning of the Ships*, 1917, by Klee and *Dreamy Improvisation*, 1913, by Kandinsky; *Dry-Cool Garden*, 1921, by Klee and *Landscape With Red Spots*, 1913, by Kandinsky; *A Hoffmannesque Tale*, 1921, by Klee and *Gentle Ascent*, 1934, by Kandinsky; *The Vase*, 1938, by Klee and *Development in Brown*, 1933, by Kandinsky.

[6]Efficiency here refers to the total payoff of the group.

[7]We ask to P3: "How many subjects out of 10, in the role of participant 2, do you expect to choose distribution LEFT?"

Table 3.1: Distributions

Distribution LEFT				
	all treatments			
P1	20			
P2	20			
P2	20			
tot	60			
Distribution RIGHT				
	Control_NI	Paintings (Cont.)	Status_L (Cont.)	Status_H (Cont.)
P1	16	Kand. = 16 (26)	Low = 16 (26)	High = 16 (26)
P2	26	Kand. = 26 (26)	Low = 26 (26)	High = 26 (26)
P3	26	Klee = 26 (16)	High = 26 (16)	Low = 26 (16)
tot	68	68	68	68

give to P1 the possibility to punish P2. Non-strategic punishment helps us to identify a social norm that links deeply a subject to a social identity, leading him even towards inefficient choices. P1 would have a reason to punish only if he perceives the behavior of P2 as unfair; hence, without identity, punishment shouldn't be a threat for P2. Instructions for the second decisional phase are handed in only at the end of the first decisional phase. Before deciding about distribution, we ask if and how much they would punish P2 in the role of P1 in both cases: if P2 chooses the LEFT and if he chooses the RIGHT distribution. Punishment is costly for P1, i.e. three points deducted from the earnings of P2 cost 1 point to P1. As in the first decisional phase we pose the direct question to P1 and P2 and we ask P3 for expectations about the punishment decision of P1. In the next screen subjects find the same decision problem about distributions as in the first decisional phase. At the end subjects uncover their roles and we present them the results of the whole experiment.

The effect of identity on distribution decisions can be detected only comparing treatment *Paintings* to a control treatment without assigning identities to the subjects. In treatment ***Control_NI*** subjects are simply P1, P2 or P3. As in the other treatments, P1 and P2 do not know their role till the end of the experiment and both has to decide in the role of P2 in the distribution choice and in the role of P1 in the punishment stage. P3 knows his role and we ask him about expectations as in the other treatments.

The two distributions differ not only in efficiency but also in inequality concerns. A subject that cares for an equal distribution among group members may choose distribution LEFT not driven by identity. To detect this portion of subjects we need a second control treatment, where P2 is not in conflict between efficiency and identity, but where he can choose between equality and efficiency. Treatment ***Control_P*** assigns the same payoffs for subjects who like the same artist in both distributions. Therefore choosing the efficient distribution RIGHT does not mean anymore a damage for the "identity partner". Subjects that here decide for the equal distribution LEFT are motivated exclusively by preferences for equality.

The last treatments ***Status_H*** and ***Status_L*** are the same as *Paintings* but assign identities using the IQ test, without the painting preference task. According to the results of the test we sort our subjects in High ability or Low ability. In the decisional phases we create group of three; in treatment *Status_H*, more able subjects will be the majority (P1=High ability, P2=High ability and P3=Low ability) and, in treatment *Status_L*, the majority is composed by the less able subjects (P1=Low ability, P2=Low ability and P3=High ability). The motivation behind these last two treatments comes from the model of Shayo (2009). He claims that identification is positively related to the status of the group and to the degree of similarity among the members. Therefore identity concerns should be stronger when the status is higher. In treatments ***Control_H*** and ***Control_L***, exactly as in *Control_P*, we change the parameters to control for equality concerns.

3.3 Theoretical Predictions

In the theoretical analysis of the "Acting White" phenomenon (Shayo 2009), it is mentioned that if subjects share a group identity, then this could lead to inefficiencies when subjects have to balance their identity with the possibility of moving socially upwards. Hypothesis H1 tests the "Acting White" phenomenon under the assumption that the generation of identity does not have an impact on inefficiency.

> **H1: In treatments where subjects have a group identity, the decision maker chooses more frequently a distribution that does not harm the ingroup subject, generating thus inefficiency.**

With hypothesis H2, we test if behavior depends on the status linked to identities, following Shayo (2009) intuition. In our experiment, the status of the group High is per definition higher than the status of the group Low, in treatments *Status_H* and *Status_L*. Meaning that inefficiency should be higher in treatment *Status_H*.

H2: The higher is the status of the group, the stronger is the group-identity effect.

Besides the argument that group identity is an explanation for "Acting White" behaviors, the literature about social identity and punishment (e.g. Fehr and Fischbacher (2004); Harris, Herrmann, and Kontoleon (2012); Hoff, Kshetramade, and Fehr (2011)) mentioned that individuals expect to get punished when they deny their identity. Hereby, punishment has not to be a direct action, but can be the exclusion from the group. We consider the punishment rationale in our experimental design by using a punishment stage in the second decisional phase. We hypothesize that "Acting White" leads to a higher degree of inefficiency when sanctions of group members are possible.

H3: Inefficiency is higher when punishment is possible.

The control treatments tests if subjects do choose LEFT because of identity concerns or because they are inequality averse. Hereby, we argue that subjects identify with their group and therefore choose more often LEFT with respect to the treatment without identity. In *Control_P*, subject do not have to balance payoff and group identity. Therefore, we should find more subjects choosing RIGHT in *Control_P*. If this is not the case, this would mean, that choosing LEFT can only be explained by inequality aversion.

3.4 Experiment's Details

We run the experiment at the Vienna Centre of Experimental Economics (VCEE) in October 2013. Each session lasted approximately one hour and in total 318 subjects, recruited via ORSEE (Greiner 2004), participated in the experiment. The software employed for programming and conduction of the experiment is z-Tree (Fischbacher 2007).

Table 3.2: Details of subjects and earnings

Treatment	subjects	average earnings € (std. dev. in brackets)
Control_NI	72	14.79 (2.31)
Paintings	48	15.09 (2.64)
Control_P	36	15.85 (2.74)
Status_L	42	15.02 (2.63)
Control_L	39	15.00 (2.21)
Status_H	42	14.67 (2.21)
Control_H	39	15.40 (2.36)
Experiment	318	15.07 (2.44)

Participants were randomly assigned to a cubicle of the computer laboratory and they were identified only by their number of seat to guarantee absolute anonymity. In the cubicle they found the instructions[8]

for the first part of the experiment. Subsequently participants were assigned to groups of three, groups did not vary for the rest of the experiment. In treatment *Control_NI*, groups were formed randomly and in the other treatments the composition of groups depended on the results of the first identity part. The group members' identities were not revealed at any time, and any form of communication was strictly prohibited.

After the first part, instructions for the first decisional phase of the second part were handed out. At this point participants discovered the composition of their team. Participant 1 and Participant 2 did not know precisely their role; we told them that they can be either Participant 1 or 2. Participant 3 knew his role. They did not receive feedbacks at the end of this phase.

Instructions for the second decisional phase were then handed out. Roles and results were shown at the end of the third phase. At the end of the experiment we summed up the performances of the phases and we pay subjects at the exchange rate of 1 point = 0.35 €.

3.5 Results

In total 318 subjects participated and earned on average 15 euros ranging from 9.80 to 23.50 euros (Table 3.2). The main task of our experiment consisted in deciding between two different distributions of

[8]Instructions differ according to the differences in treatment.

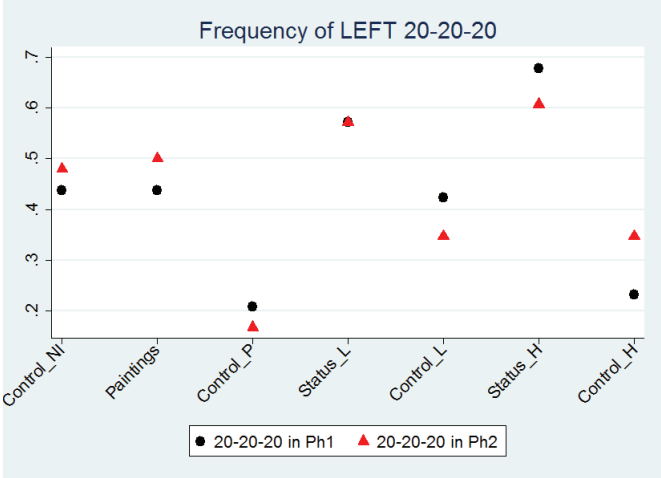

Figure 3.2: Frequency of LEFT choices. Black dots refer to data of the first decisional phase and red triangle to the second decisional phase.

wealth for the own group, as explained in Section 3.2. In Figure 3.2 we can see the proportion of subjects that chose distribution LEFT sorted by treatments. Black dots represent data of the first decisional phase and red squares refer to the second decisional phase were punishment is allowed.

3.5.1 Minimal Group Paradigm

First of all we compare the control treatment without identity *Control_NI* with the two treatments where identity is created by preferences over paintings, *Paintings* and *Control_P*. In *Control_NI* already the 43.75% choose the inefficient distribution LEFT that gives 20 points to each member of the group.[9] In our setting this percentage represents the proportion of subjects that prefers equality not maximizing the personal payoff and forgoing efficiency.

In treatment *Paintings* the percentage of LEFT choices does not vary at all (43.75% - Fisher's exact test p=1.000). Therefore identity created in the laboratory through preferences over painting is not strong enough to increase inefficiency. Another possible explanation is that using the strategy method also in *Control_NI* we are already creating a kind of identity among the two subjects that have some decisional power (P1

[9]Experiments using Dictator Games find from 20% to 40% of subjects that split the endowment evenly (Andreoni and Miller 1995; Forsythe et al. 1994)

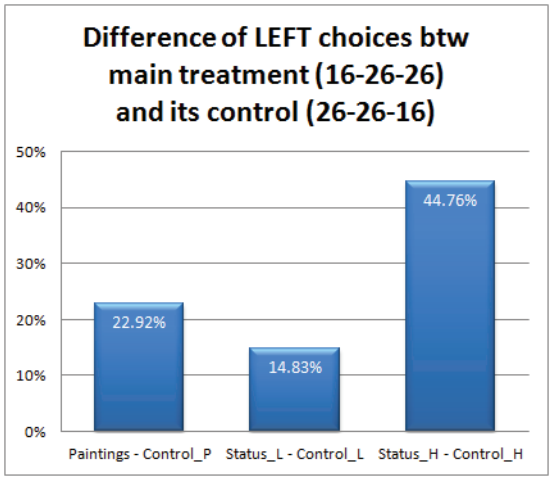

Figure 3.3: Difference of LEFT choices' frequency between each main treatment and its control.

and P2).

When instead the efficient/unequal choice RIGHT benefits both members of the group, as in *Control_P* treatment, identity leads participants to decrease substantially inefficiency with only the 20.83% of LEFT choices (Fisher's exact test p=0.065 *Paintings* vs. *Control_I* and p=0.048 *Control_NI* vs. *Control_P*).

The effect of identity becomes really clear comparing *Paintings* and *Control_P*. Inefficiency increases of 23% when efficiency implies a disadvantage for the group mate. Figure 3.3 shows the percentage difference of LEFT choices between the main treatments (where P2 is in conflict between identity and efficiency) and the relative control treatments where efficiency means also helping the group mate.

3.5.2 Status

Following the intuition of Shayo (2009), we assigned identities to the groups varying also their status. Using the results of an IQ test we created groups, in treatment *Status_L*, where P1 and P2 were less able and P3 had a higher IQ and the opposite situation in treatment *Status_H*.

As the model cited predicts, identity is stronger when the status is higher. In treatment *Status_L*, the 57.14% of subjects chose the inefficient distribution LEFT, in treatment *Status_H* the frequency of inefficiency increased till 67.84%, as shown in Figure 3.2 (Fisher's exact test p=0.187

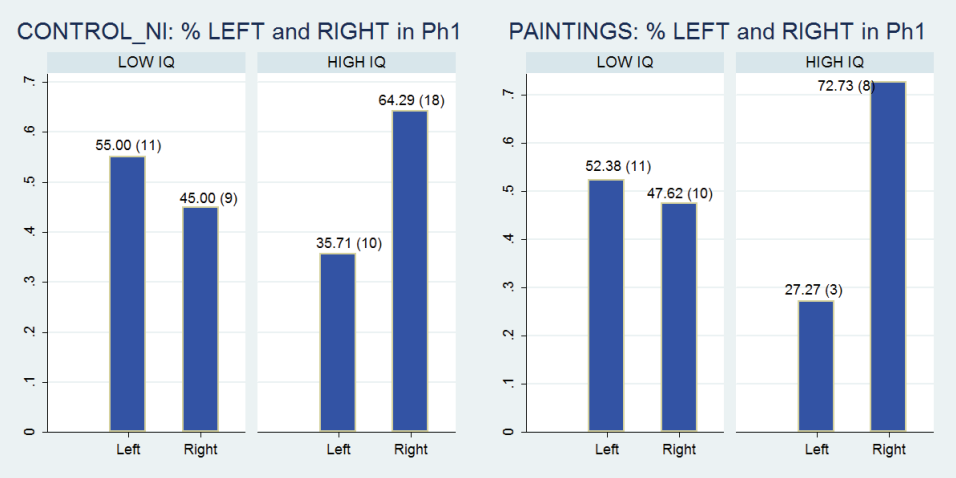

Figure 3.4: Percentage of LEFT and RIGHT choices (number of observations in brackets) sorted by the performance of the IQ test in treatments *Control_NI* and *Paintings*. Subjects below the median performance are named Low_IQ and subjects above are in High_IQ category.

Control_NI vs. *Status_L* and p=0.036 *Control_NI* vs. *Status_H*).

Another way to see this effect is to look at the difference of LEFT choices between the treatment where P2 is in conflict (identity vs. efficiency) and its control where efficiency means also helping the own group. In Figure 3.3 we can see that when efficiency (RIGHT) hurts the group mate, the fraction of inefficient choices (LEFT) increases of about 15% in *Status_L* and even of about 45% when the status is high.

Therefore in our setting, where identity yields inefficiency, a higher status linked to the social identity worsens the inefficiency problem.

Identity in our *Status* treatments is created through the result of an IQ test. A natural concern therefore could be selection. It could be that high skilled subjects usually choose more often the LEFT choice maybe driven by some equality concerns. Given that the IQ test was performed also in treatments *Control_NI* and *Paintings*, we can now split ex-post the sample of these treatments between High and Low as we did for the *Status* treatments. In Figure 3.4 we can see this categorization. In the left part we find data of the *Control_NI* treatment and in the right part of the *Paintings* treatment. We grouped subjects that performed below the median in the Low_IQ category and the better in the High_IQ one. The labels LEFT and RIGHT on the horizontal axis identify the two

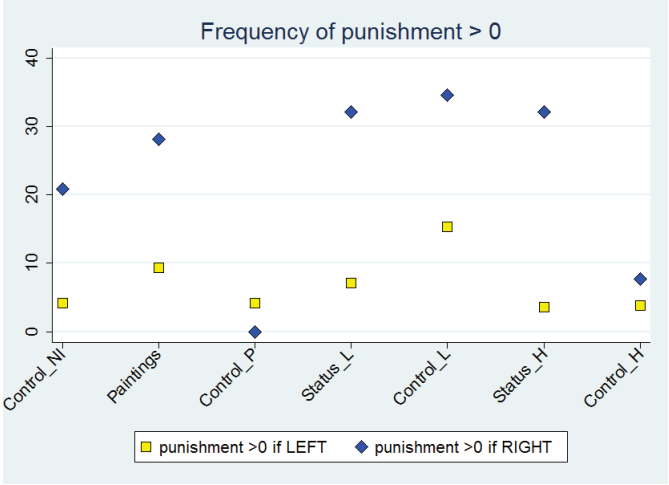

Figure 3.5: Frequency of punishment.

distributions that the subjects can choose.

From the graph we can see that low skilled subjects are more or less evenly split between the two distributions, but more interesting subjects of the High_IQ category chose more often the efficient/unequal distribution RIGHT (Fisher's exact test p=0.054 LEFT vs. RIGHT for High_IQ category.). High_IQ subjects tend to opt for efficiency and maximization of their payoffs when identity does not matter, but when they are grouped together and they face the trade-off between identity betrayal and efficiency, they completely reverse the preference picking the LEFT distribution even at a high personal cost (20 points instead of 26).

3.5.3 Punishment

In the experimental literature, punishment is often used to detect the presence of a social norm, relying on the characteristic that punishment is costly also for the subjects that send it (Harris, Herrmann, and Kontoleon 2012).

In our experiment punishment is introduced in the second decisional phase. We have seen that identity is a strong deterrent to hurt the ingroup, therefore our hypothesis H3 suggests that inefficiency should increase in case the group mate has the possibility to punish.

Figure 3.5 shows the percentage of subjects, in the role of P1, that are willing to punish their group mate.[10] On the horizontal axis we find

[10]In the experiment, subjects has 5 possible choice of punishment: 0 at cost of 0, 3

all the treatments, the yellow squares represent the frequency of punishment if P2 chooses LEFT (inefficient/equal) and the blue diamonds if P2 chooses RIGHT (efficient/unequal).

First of all we note that in the main treatments (not the control ones) there is a significative difference between LEFT and RIGHT.[11] In these treatments subjects tend to punish more the choice of RIGHT; this means that even if RIGHT creates an overall higher wealth, it deserves to be punished by the group mate because it damages the ingroup. In the control treatments *Control_P* and *Control_H*, the punishment for RIGHT drops a lot, since now the group mate receives an advantage from this choice.[12] If the reason for punishment would have derived from inequality aversion, we would have seen positive punishment in case of RIGHT also in the control treatments; therefore we can rule out this hypothesis and motivate the behaviors with identity concerns.

Looking at the punishment we can see that there is a social norm that drives subjects to punish choices that damage the ingroup. However, the level of punishment does not vary with the strength of identity.

Hypothesis H3 claims that the frequency of LEFT choices may increase when the group mate P1 has the possibility to punish the P2. For detecting the effect of punishment on behaviors we have to look back at Figure 3.2. Data from the first decisional phase without punishment are represented by the black dots, and from the second decisional phase with punishment by the red triangles. In all treatments the frequency of LEFT choices does not change between phases.[13] Therefore the threat of punishment does not influence behaviors.

3.6 Conclusion

"Acting White" is the expression used to identify situations where the link to a social identity brings to underinvestment and inefficiency. In our work we replicate this phenomenon with a laboratory experiment.

at cost of 1, 6 at cost of 2, 9 at cost of 3 and 12 at cost of 4. In the data analysis we create only two categories out of these data: $punishment = 0$ and $punishment > 0$.

[11] Wilcoxon rank-sum tests: *Control_NI* p=0.011, *Paintings* p=0.062, *Status_L* p=0.015, *Status_H* p=0.001

[12] In treatment *Control_L* the level of punishment is not different for the two distributions as in *Control_P* and *Control_H*; however the punishment in case of RIGHT is significative higher and similar to the level of the main treatments. We don't have a clear explanation for this evidence yet.

[13] McNemar test: *Control_NI* p=0.754, *Paintings* p=0.687, *Control_P* p=1.00, *Status_L* p=1.00, *Control_L* p=0.687, *Status_H* p=0.625, *Control_H* p=0.508.

We created groups of three subjects where two of them were linked by a social identity and they had to decide how to distribute wealth among the members of in and the out-group. To see how identity influences behaviors we created a conflict between identity, own income and total income of the group.

In the context where increasing the own payoffs generates a positive externality for the outgroup but a negative externality for the ingroup, our data highlight that social identity holds back individual investments when this means hurting the social identity partner. A related result can be found in the paper of Hadnes, Vollan, and Kosfeld (2012). They run a real-effort experiment with tailors in Burkina Faso. They saw that when workers were linked to a solidarity group that shares part of the income, the individual productivity drops. Therefore also in their study we observe that in some contexts the link to some social group can generate inefficiency in the society.

The level of inefficiency is also linked to the status of the social group. Subjects feel a deeper attachment to their social identity when its status is perceived as higher. This emotional state leads subjects to choose more often the distribution that does not hurt the ingroup, increasing thus inefficiency in our setting.

The effect of identity on inefficiency is not driven by the fear of punishment form the identity partner. Punishment does not alter the behaviors, however through the amount of punishment inflicted we can see that there is a social norm that favors loyalty towards the social identity group to the detriment of efficiency.

A natural extension to this work is to find ways that can alleviate the inefficiency problem; compensation could be one of those. Coming back to the education example, when the black community keep its valuable members from investing on education, it is depriving the whole society of a potentially assets. The rest of the society could therefore find some compensation tools that reduce ostracism of the identity group.

Bibliography

Ainsworth-Darnell, James W. and Douglas B. Downey (1998). "Assessing the oppositional culture explanation for racial/ethnic differences in school performance". In: *American Sociological Review*, pp. 536–553.

Akerlof, George A. and Rachel E. Kranton (2000). "Economics And Identity". In: *The Quarterly Journal of Economics* 115.3, pp. 715–753.

Albert, Max et al. (2007). "Are we nice(r) to nice(r) people? An experimental analysis". In: *Experimental Economics* 10.1, pp. 53 –69.

Andreoni, James and John H Miller (1995). "Auctions with artificial adaptive agents". In: *Games and Economic Behavior* 10.1, pp. 39–64.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul (2012). *Team Incentives: Evidence from a Firm Level Experiment*. IZA Discussion Papers 6279. Institute for the Study of Labor (IZA).

Benabou, Roland and Jean Tirole (2003). "Intrinsic and extrinsic motivation". In: *The Review of Economic Studies* 70.3, pp. 489–520.

— (2006). "Incentives and Prosocial Behavior". In: *American Economic Review* 96.5, pp. 1652–1678.

Berg, Joyce, John Dickhaut, and Kevin McCabe (1995). "Trust, Reciprocity, and Social History". In: *Games and Economic Behavior* 10.1, pp. 122 –142.

Bernheim, B Douglas (1994). "A Theory of Conformity". In: *Journal of Political Economy* 102.5, pp. 841–77.

Bleeke, Joel and David Ernst (1991). "The way to win in cross-border alliances". In: *Harvard Business Review* 69.6, pp. 127–135.

Bolton, Gary E and Axel Ockenfels (2000). "ERC: A Theory of Equity, Reciprocity, and Competition". In: *The American Economic Review* 90.1, pp. 166–193.

Cason, T.N. and L. Gangadharan (2010). *Cooperation Spillovers and Price Competition in Experimental Markets*. working paper.

Charness, Gary (2012). *Efficiency, Team building, and Spillover in a Public-goods Game*. University of California at Santa Barbara, Eco-

nomics Working Papern Series qt2np178xh. Department of Economics, UC Santa Barbara.

Chen, Yan and Sherry Xin Li (2009). "Group Identity and Social Preferences". In: *The American Economic Review* 99.1, pp. 431–457.

Cherry, Todd L., Thomas D. Crocker, and Jason F. Shogren (2003). "Rationality spillovers". In: *Journal of Environmental Economics and Management* 45.1, pp. 63 –84.

Cherry, Todd L. and Jason F. Shogren (2007). "Rationality crossovers". In: *Journal of Economic Psychology* 28.2, pp. 261 –277.

Cook, Philip J. and Jens Ludwig (1997). "Weighing the "Burden of 'Acting White'": Are There Race Differences in Attitudes toward Education?" In: *Journal of Policy Analysis and Management* 16.2, pp. 256–278.

Corneo, Giacomo G. (1997). "The theory of the open shop trade union reconsidered". In: *Labour Economics* 4.1, pp. 71–84.

Cox, James C., Daniel Friedman, and Steven Gjerstad (2007). "A tractable model of reciprocity and fairness". In: *Games and Economic Behavior* 59.1, pp. 17 –45.

Crawford, Vincent P. (1979). "A Procedure for Generating Pareto-Efficient Egalitarian-Equivalent Allocations". In: *Econometrica* 47.1, pp. 49–60.

Currarini, Sergio, Matthew O Jackson, and Paolo Pin (2009). "An economic model of friendship: Homophily, minorities, and segregation". In: *Econometrica* 77.4, pp. 1003–1045.

Fehr, Ernst (2009). "On the Economics and Biology of Trust". In: *Journal of the European Economic Association* 7.2-3, pp. 235–266.

Fehr, Ernst and Urs Fischbacher (2004). "Third-party punishment and social norms". In: *Evolution and human behavior* 25.2, pp. 63–87.

Fehr, Ernst and Klaus M. Schmidt (1999). "A Theory Of Fairness, Competition, and Cooperation*". In: *Quarterly Journal of Economics* 114.3, pp. 817–868.

Fischbacher, Urs (2007). "z-Tree: Zurich toolbox for ready-made economic experiments". In: *Experimental Economics* 10.2, pp. 171–178.

Fordham, Signithia and John U. Ogbu (1986). "Black students' school success: Coping with the "burden of 'acting white'"". In: *The Urban Review* 18.3, pp. 176–206.

Forsythe, Robert et al. (1994). "Fairness in Simple Bargaining Experiments". In: *Games and Economic Behavior* 6.3, pp. 347–369.

Frey, Bruno S (1992). "Tertium datur: Pricing, regulating and intrinsic motivation". In: *Kyklos* 45.2, pp. 161–184.

Frey, Bruno S and Reto Jegen (2001). "Motivation crowding theory". In: *Journal of economic surveys* 15.5, pp. 589–611.

Fryer, Roland G. (2006). ""Acting White"". In: *Education Next* 6, pp. 52–59.

Fryer, Roland G. and Paul Torelli (2010). "An empirical analysis of 'acting white'". In: *Journal of Public Economics* 94.5-6, pp. 380–396.

Gneezy, Uri, Stephan Meier, and Pedro Rey-Biel (2011). "When and why incentives (don't) work to modify behavior". In: *The Journal of Economic Perspectives* 25.4, pp. 191–209.

Gneezy, Uri and Aldo Rustichini (2000). "Pay enough or don't pay at all". In: *The Quarterly Journal of Economics* 115.3, pp. 791–810.

Greiner, Ben (2004). *The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in Economics*. Working Paper Series in Economics 10. University of Cologne, Department of Economics.

Grogger, Jeffrey and Gordon H Hanson (2011). "Income maximization and the selection and sorting of international migrants". In: *Journal of Development Economics* 95.1, pp. 42–57.

Grossman, Sanford J. and Oliver D. Hart (1986). "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration". In: *Journal of Political Economy* 94.4, pp. 691–719.

Guimond, Serge, Stephane Dif, and Annabelle Aupy (2002). "Social identity, relative group status and intergroup attitudes: When favourable outcomes change intergroup relations... for the worse". In: *European journal of social psychology* 32.6, pp. 739–760.

Hadnes, Myriam, Björn Vollan, and Michael Kosfeld (2012). "The dark side of solidarity". In: *Unpublished manuscript, Goethe-University Frankfurt*.

Hamilton, Barton H, Jack A Nickerson, and Hideo Owan (2003). "Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation". In: *Journal of political Economy* 111.3, pp. 465–497.

Harris, Donna, Benedikt Herrmann, and Andreas Kontoleon (2012). *When to favour your own group? : the threats of costly punishments and in-group favouritism*. Department of Economics discussion paper series / University of Oxford 628. Oxford : Univ. of Oxford, Dep. of Economics.

Hoff, Karla, Mayuresh Kshetramade, and Ernst Fehr (2011). "Caste and Punishment: the Legacy of Caste Culture in Norm Enforcement". In: *Economic Journal* 121.556, F449–F475.

Klor, Esteban F. and Moses Shayo (2010). "Social identity and preferences over redistribution". In: *Journal of Public Economics* 94.3-4, pp. 269–278.

Knez, Marc and Colin Camerer (2000). "Increasing Cooperation in Prisoner's Dilemmas by Establishing a Precedent of Efficiency in Coordination Games". In: *Organizational Behavior and Human Decision Processes* 82.2, pp. 194 –216.

Leonard, Nancy H, Laura L Beauvais, and Richard W Scholl (1995). "A self concept-base model of work motivation". In: *Academy of Management Proceedings*. Vol. 1995. 1. Academy of Management, pp. 322–326.

Liberman, Varda, Steven M. Samuels, and Lee Ross (2004). "The Name of the Game: Predictive Power of Reputations versus Situational Labels in Determining Prisoner's Dilemma Game Moves". In: *Personality and Social Psychology Bulletin* 30.9, pp. 1175 –1185.

Mengel, F. and E. Sciubba (2010). *Extrapolation and Structural Learning in Games*. working paper. Maastricht University.

Naef, Michael and Jürgen Schupp (2009). *Measuring Trust: Experiments and Surveys in Contrast and Combination*. IZA Discussion Papers 4087. Institute for the Study of Labor (IZA).

Nagel, Rosemarie (1995). "Unraveling in Guessing Games: An Experimental Study". In: *American Economic Review* 85.5, pp. 1313 –1326.

Pilon, Mary (2010). "I Love You, You're Perfect, Now Sign Here". In: *The Wall Street Journal*.

Rainer, Helmut (2007). "Should we write prenuptial contracts?" In: *European Economic Review* 51.2, pp. 337 –363.

Roccas, Sonia (2003). "The effects of status on identification with multiple groups". In: *European Journal of Social Psychology* 33.3, pp. 351–366.

Rosso, Brent D, Kathryn H Dekas, and Amy Wrzesniewski (2010). "On the meaning of work: A theoretical integration and review". In: *Research in Organizational Behavior* 30, pp. 91–127.

Ryan, Richard M and Edward L Deci (2000). "Intrinsic and extrinsic motivations: Classic definitions and new directions". In: *Contemporary educational psychology* 25.1, pp. 5–67.

Seabright, Paul (2004). *Continuous Preferences Can Cause Discontinuous Choices: An Application to the Impact of Incentives on Altruism*. CEPR Discussion Papers 4322. C.E.P.R. Discussion Papers.

Segal, Carmit (2012). "Working when no one is watching: Motivation, test scores, and economic success". In: *Management Science* 58.8, pp. 1438–1457.

Shayo, Moses (2009). "A model of social identity with an application to political economy: Nation, class, and redistribution". In: *American Political Science Review* 103.2, pp. 147–174.

Sheldon, Kennon M et al. (2003). "Applying self-determination theory to organizational research". In: *Research in personnel and human resources management* 22, pp. 357–393.

Shotter, Andrew (1999). "Worker trust, system vulnerability, and the performance of work groups". In: *Economics, values, and organization*. Cambridge University Press, pp. 364–407.

Spence, Michael (1973). "Job market signaling". In: *The quarterly journal of Economics* 87.3, pp. 355–374.

Tajfel, Henri and John Turner (1979). "An Integrative Theory of Intergroup Conflict". In: *The Social Psychology of Intergroup Relations*, pp. 33–47.

Tajfel, Henri et al. (1971). "Social categorization and intergroup behaviour". In: *European journal of social psychology* 1.2, pp. 149–178.