Give and Take

Three Essays on Giving behind the Veil of Ignorance,

Taking with Deterrent Incentives, and Educational Attainment

Inaugural-Dissertation
zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Ludwig-Maximilians-Universität München

2008

vorgelegt von

Hannah Hörisch

Referent:	Prof. Dr. Klaus M. Schmidt
Korreferent:	Prof. Dr. Joachim Winter
Promotionsabschlussberatung:	16. Juli 2008

Acknowledgements

First and foremost, I would like to thank my supervisor Klaus Schmidt for all his critical questions and constructive comments that have shaped this thesis. I am also very grateful for his general support and encouragement as well as many inspiring and challenging discussions.

I have always enjoyed working at the Seminar of Economic Theory at the University of Munich – an environment that was formed by my colleagues at the chair. Thank you all for very pleasurable collaboration in excellent atmosphere. Special thanks go to my co-author Christina Strassmair. I have enjoyed working with her a lot and have profited enormously from our discussions.

I have received insightful and helpful comments from too many colleagues in Munich and at many conferences and seminar presentations to list them all – still, I am indebted to all of them. I thank Christian Traxler, Christoph Bauner, Felix Hörisch, Georg Gebhardt, Guido Schwedt, Joachim Winter, Ludger Wößmann, Maria Lehner, Martin Kocher, Matthias Sutter, Nadine Riedel, Sandra Ludwig and Simone Kohnz for their comments and time. Julia Bersch and Florian Kajuth had always time to listen to my doubts and to help with their remarks.

I also thank the Stockholm School of Economics, the local Ph.D. students and especially Magnus Johannesson for hosting me as a visiting Ph.D. scholar in Stockholm – a stay that I have enjoyed and profited from a lot.

Financial support from the Deutsche Forschungsgemeinschaft via SFB Transregio 15 is gratefully acknowledged.

Last but by no means least, I thank my boyfriend Björn and my family for their love and constant support. This is for you.

Hannah Hörisch

Contents

List of Tables

List of Figures

Preface

This dissertation consists of three self-contained chapters that are contributions to research in experimental, behavioral and applied empirical economics. Each chapter has its own introduction and appendix and can be read independently of the other two chapters. Still, to some extent, the three chapters can be subsumed under the common theme of "give and take".

The first chapter questions the utilitarians' claim that acting according to maximin preferences behind the veil of ignorance is only optimal for infinitively risk averse individuals. It shows experimentally that maximin preferences are compatible with any degree of risk aversion if social preferences for equality are sufficiently strong. In the experiment subjects play variants of the dictator game without and behind the veil of ignorance. A standard dictator game consists of the decision how much (to take from the own initial endowment and) to give to another subject. Thus the first chapter investigates giving behavior.

Taking or stealing is the central theme of the second chapter. Experimental subjects have the possibility to steal from another subject's initial endowment. By varying the intensity of deterrent incentives (i.e. detection probability and fine for stealing) the experiment tests Becker's (1968) deterrence hypothesis that crime rates are decreasing in deterrent incentives.

Finally, the third chapter uses field data to investigate whether and how parental employment affects children's educational attainment. Parents' decision to participate in the labor market largely determines how much time and good inputs they can give to their children. Economic models of knowledge acquisition in childhood usually assume that children take the time and good inputs and convert them into educational achievement.

The first two chapters of this dissertation use experimental methods. In 2002, Vernon Smith, one of the earliest experimenters in economics, has been awarded the Nobel price in economics. This manifests that experimental methods have become an established tool in empirical economic analysis. One very fundamental consequence of experimental evidence has been to question the pure self-interest assumption underlying standard neo-classical economic analysis. By now, it is widely accepted that preferences are heterogeneous: while some individuals' preferences are consistent with pure self-interest others have other-regarding preferences. Many different concepts of other-regarding preferences have been developed: in addition to own material well-being preferences are assumed to potentially depend on other individuals' absolute or relative material well-being, other individuals' fair or unfair behavior or other individuals' type such as generally being selfish, spiteful or altruistic.[1] This dissertation uses concepts of and finds evidence for models of social preferences. The utility function of an individual with social preferences does not only depend on his own material well-being but may also be a function of the allocation of resources within his reference group. In particular, we use models of inequity aversion as proposed by Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) in which people are assumed to suffer from unequal allocations.

Outcomes in competitive experimental markets often converge to equilibrium predictions of standard neo-classical theory and thus may be unaffected by the existence of social preferences. In contrast, social preferences shape the outcomes in distributional games such as dictator and ultimatum games.[2] The first two chapters of this dissertation investigate distributional games in non-competitive experimental environments. Hence, we expect to find and indeed do find that players with social preferences crucially influence our results.

[1] For a recent survey on theories of other-regarding preferences see Fehr and Schmidt (2006, Section 3).

[2] Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) show that both empirical findings in competitive markets and distributional games are consistent with their model of heterogeneous - inequity-averse and selfish - types of players. The intuition is that there exist important interactions between the distribution of preferences and the strategic environment: only if single players can affect relative material payoffs, equality considerations will affect experimental outcomes. This is usually not the case in competitive environments but in distributional decisions.

The classic game used in experimental economics to elicit a player's distributional preferences is the dictator game. A dictator game is a very simple two-player game in which the first player, the dictator, proposes a split of a given pie. The second player, the receiver, is completely passive. Both players are paid according to the dictator's proposal. Since there is no strategic interaction between the two players, the dictator's proposal reflects his pure distributional preferences. The dictator game has been introduced by Kahneman, Knetsch and Thaler (1986) and Forsythe et al. (1994). Camerer (2003, Ch. 2) surveys the literature on dictator games. The experiments presented in chapters 1 and 2 are inspired by the dictator game, but heavily modify it.

The game played in chapter 1 is a dictator game with two additional features: first, our dictator game is characterized by an efficiency loss of 50 % for each unit that the decision maker assigns the receiver. The efficiency loss introduces a trade-off between equality and efficiency. Andreoni and Vesterlund (2001), Andreoni and Miller (2002) and Fisman, Kariv and Markovits (2007) also use dictator games with an efficiency loss. Second, we add role uncertainty to implement the veil of ignorance, i.e. at the time the transfer decision is made the decision maker does not yet know whether he will be paid as dictator or receiver.

The second chapter analyzes behavior in a take game, the mirror image of the standard dictator game. In a take game the dictator decides how much to take from the positive initial endowment of the second passive player (instead of how much to give to the second player as in a standard dictator game). Bardsley (2005), List (2005) and Krupka and Weber (2006) use games in which dictators can choose between giving to or taking from the second passive player. Since our aim is to simulate stealing in the laboratory, we, in contrast, use a pure take game frame. Furthermore, we augment the take game with varying levels of incentives that make taking less attractive since we aim at testing the deterrence hypothesis.

The experiment presented in the first chapter questions the view that behind the veil of ignorance maximin preferences necessarily represent preferences with infinite risk aversion. The philosopher John Rawls (1971) argued that behind the veil of ignorance people would vote for the difference principle, i.e. in favor of maximizing the utility of the worst off individual. Utilitarians generally do not accept this claim and stress that voting for the difference principle would only be strictly optimal for infinitely risk averse individuals. In contrast, we hypothesize that maximin preferences are compatible with any degree of risk aversion if social preferences for equality are sufficiently strong. We test this hypothesis experimentally.

The experimental design is based on a dictator game with two additional characteristics. First, we implement the veil of ignorance by introducing role uncertainty: each subject decides how many units of a given pie the dictator will give away to the receiver *before* the subject is randomly assigned the role of dictator or receiver with equal probability. Each subject will be paid in his assigned role according to his own decision how many units the dictator will transfer to the receiver. Implementing the veil of ignorance removes the possibility to favor oneself over the other subject and introduces risk. Second, there is an efficiency loss of 50 % for units that are transferred from the dictator to the receiver. The efficiency loss makes insurance costly and hence allows measuring a subject's degree of risk aversion. Additionally, it introduces a trade off between equality and efficiency in a two person game.

The core of the analysis is the comparison of the following two treatments: the veil of ignorance treatment that is a dictator game with efficiency loss and role uncertainty as described above and the risk treatment. The risk treatment is identical to the veil of ignorance treatment except for one difference. It is a one-person game: each subject decides how to allocate the pie across the states of being dictator or being receiver and is randomly assigned the position of either dictator or receiver afterwards. However, the position not assigned to the decision maker is not filled in by a second subject. The two treatments are completely identical in terms of risk, but social preferences can only matter if the decision affects at least one other subject besides the decision maker, i.e. in the veil of ignorance treatment, but not in the risk treatment. Any difference between the risk and the veil of ignorance treatment must be induced by social preferences. In contrast to other economic experiments that implement the veil of ignorance in the laboratory our experimental design allows separating the effects of risk aversion and social preferences behind the veil of ignorance.

We find that behind the veil of ignorance only a minority of subjects opts for the difference principle. Decisions in the risk and the veil of ignorance treatment do not differ significantly for men. In contrast, behind the veil of ignorance social preferences are a second significant motivation besides risk for women and induce a stronger concern for equality. Thus, our results for women imply that voting for the difference principle is not only optimal for infinitely risk averse individuals, but also for individuals with strong social preferences for equality. Furthermore, our results contribute to the growing literature on gender differences in social preferences and risk attitudes.

The experimental game in the second chapter focuses no longer on a pure distributional decision (under risk), but on how exogenously set incentives influence a dictator's distributional decision. Furthermore, we switch from a give to a take frame of the dictator game.

How to effectively combat crime is the topic of an ongoing and vivid debate in the general public, among politicians, judges, social workers and criminologists. Becker's (1968) deterrence hypothesis is the key contribution of economists to this debate. It relies on the power of pure incentives and states that crime rates are weakly decreasing in deterrent incentives, i.e. a crime's detection probability and the level of punishment. Empirical evidence from field data is often - but far from always - consistent with the deterrence hypothesis and shows that variations in deterrent incentives can only explain a small part of the variation in crime (see Glaeser, 1999 and Eide, 2000). A serious drawback of field data tests of the deterrence hypothesis is the lack of appropriate data. For example, individual level data are rarely available and results based on aggregate data may suffer from simultaneity bias. We are the first to explicitly test the deterrence hypothesis in the laboratory which permits to exogenously vary deterrent incentives and to obtain representative individual level data. We ask a very basic but important question: do deterrent incentives work?

In our take game design, two subjects, A and B, are matched randomly. Subject A is passive and has a higher initial endowment than subject B. Subject B can decide how much to take away (steal) from A's initial endowment. With probability $1-p$, this amount is transferred from A to B. With probability p ("detection probability"), however, this amount is not transferred and a fixed fine is deducted from B's initial endowment if B attempted to take a strictly positive amount. We conduct six different treatments in which we vary detection probability and fine. The combined effect of detection probability and fine ranges from zero to strong deterrent incentives, i.e. levels of incentives such that taking subject A's whole initial endowment does not pay off in expectation. Since weak deterrent incentives are especially relevant in real life, four out of six treatments use small or intermediate incentives, i.e. incentives such that taking everything pays off in expectation. Each subject participates in two different treatments sequentially such that we can analyze taking behavior both across subjects and for a given subject.

For most experimental sessions, we use neutrally framed instructions because our primary aim is to test the economic approach to crime that relies on incentive effects only. As a robustness check we run a few "morally framed" sessions in which we talk about "stealing" instead of "transfer decisions" and about a "fine" instead of "minus points".

Our neutrally framed across subjects results clearly reject the deterrence hypothesis. We find that incentives backfire: on average subjects take significantly more in treatments with intermediate deterrent incentives than in the absence of incentives. Only very strong incentives deter. Both our neutrally framed across and within subjects results can be explained by a model of two types: about 50 % of our subjects are selfish and react to deterrent incentives as predicted by the deterrence hypothesis. The other 50 % are fair-minded subjects for whom deterrent incentives backfire by crowding out fairness concerns. In our morally framed sessions we observe backfiring of incentives, too. Furthermore, we find that detection probability and fine seem to be interchangeable instruments.

Finally, the last chapter uses field data to investigate whether and how parental employment affects children's educational attainment. This is an important question since it concerns the organization of every day life of virtually all families and has far-reaching policy implications for the public provision of child care facilities, maternity leave legislation or the design of welfare and tax systems. Though there are numerous studies on the effect of parental employment on children's outcomes, the existing evidence is very inconclusive. It stems nearly exclusively from U.S. and British data, focuses largely on maternal employment and children's outcomes at young ages. Chapter 3 adds to the existing literature along all these dimensions: it uses a large German panel data set, the German Socioeconomic Panel, for the first time to address the effect of both maternal and paternal employment on children's educational attainment. The dependent variable of our analysis is attendance of high secondary school track that has been shown to be an important predictor of later labor market success (Dustmann, 2004). Most importantly, in contrast to the vast majority of existing studies we explicitly address potential endogeneity problems: first, to deal with selection of parents in the labor market we estimate a model on sibling differences that controls for all unobserved time-invariant parent and household characteristics. Second, we tackle a potential reversed causality problem, i.e. the fact that parents' decision to work may be affected by their child's ability which in turn partially determines educational success). To do so we only use parental employment when children are aged 0-3 such that their ability is not yet fully revealed, exclude disabled children from the analysis and use parental years of education as a proxy for children's ability.

We analyze two potential effects of parental employment: a positive effect through higher household income and a negative time effect since working parents spent less time with their children. We use various direct and indirect measurements of parent's time inputs in raising their children when children are aged 0-3: daily time spent on child care, weekly hours worked and the number of years in which parents worked full time, part time or not at all.

In sum, we neither find evidence for a positive income effect nor a negative time effect. Controlling for household income we can statistically rule out that having a mother who works one hour more per week lowers the probability of high secondary track attendance by more than 0.1 percentage points, an economically negligible number. Actually, all coefficients of maternal employment are positive, but not significant at conventional levels (though at a 9 to 11 % significance level). Coefficients of father's employment are precisely estimated but too small to be significant. When we use parental time spent on child care instead of parental employment information as key explanatory variable, coefficients are also too small to be significant. Taken together, our results imply that it is not parental employment or quantity of parent-child interactions that is decisive for children's educational attainments, but, for example, birth order within a family, age relative to classmates or parental characteristics like their level of education.

Chapter 1

Is the veil of ignorance only a concept about risk? An experiment

1.1 Introduction

Our experiment explores the relationship between social preferences and Rawls' difference principle that economists have formalized by maximin preferences. In his book "A Theory of Justice" (1971) the philosopher John Rawls coined the term "veil of ignorance" for the following thought experiment: Behind the veil of ignorance, nobody knows which future position in society he (as well as other individuals) will be assigned when deciding how to distribute resources across different positions. According to Rawls society would agree behind the veil of ignorance that the difference principle should constitute the basis of the social contract. The difference principle states that society should maximize the utility of the individual that is worst off. Utilitarians have asserted that being in favor of the difference principle is only strictly optimal for infinitely risk averse individuals and thus, have dismissed the difference principle and maximin preferences as unrealistic. However, the Utilitarian's argument assumes that everybody is only interested in his own material payoff. In contrast, theories on social preferences assume that people are self-interested to some degree, but also care about (the payoffs of) others.[3] In this paper, we argue that if people have social preferences,

[3] Focusing on the distribution of payoffs the notion of social preferences we use is most closely related to Fehr and Schmidt's (1999) and Bolton and Ockenfels' (2000) models of inequity aversion.

they could be in favor of an egalitarian distribution even if they are risk neutral.

Our experiment implements the veil of ignorance in the laboratory[4] and tests whether decisions behind the veil of ignorance are only driven by risk attitudes or also by social preferences. Assume decisions behind the veil of ignorance reflect (impartial) social preferences for equality in addition to risk aversion. Then the difference principle is consistent with any degree of risk aversion as long as social preferences for equality are sufficiently strong to make individuals opt for a completely equal distribution.

Implementing the veil of ignorance we measure social preferences that are free of self-interest in a narrow sense ("impartial social preferences"). In other words, impartial social preferences are an individual's preferences over distributions of payoffs to himself and his reference group when favoring oneself over the others is not possible. Information on people's impartial social preferences can be useful for many aspects of policy design, e.g. the design of tax, social security or public health insurance systems. Imagine, as an example for eliciting social preferences, a survey in which you ask a poor person whether he is in favor of more redistribution. If you get the answer "yes" you cannot interpret it unambiguously: does this person prefer more redistribution because he is likely to profit from it? Or does this person have an innate preference for a more equal society? In contrast, if you had asked this person behind the veil of ignorance and had received the (now impartial) answer "yes" you would have known that the latter is true (or that this person is risk averse).

Our experiment uses a three treatment design: the *dictator game treatment* is a dictator game with a 50 % efficiency loss. A dictator game is a two player game in which the first player, the dictator, proposes a split of a given pie. The second player, the receiver, is passive. Both players are paid according to the dictator's proposal.[5] Our specific variant of the dictator game is characterized by an efficiency loss of 50 % for units that are transferred from the dictator to the receiver. Consequently, a

For a recent survey on the literature on social preferences see Fehr and Schmidt (2006).

[4] With any implementation of the veil of ignorance in the laboratory subjects will know much more than in Rawls' original position, e.g. they will know their sex and ability. Still, the implementation of the veil of ignorance is perfect with respect to subjects' positions and implied payoffs. Hence, in our experimental setup we can measure subjects' risk attitudes and potential social preferences behind the veil of ignorance and this is what we aim at.

[5] In this variant, the dictator game was first introduced by Forsythe et al. (1994).

trade off between equality and efficiency[6] arises: a more equal allocation can only be achieved by transferring more which in turn induces a larger efficiency loss. Our second treatment, the *veil of ignorance treatment*, is characterized by the same efficiency loss, but adds role uncertainty to implement the veil of ignorance: each participant decides how many units of a 12 unit pie the dictator will give away to the receiver *before* he is assigned the role of dictator or receiver with equal probability. Finally, each participant will be paid according to his own choice how many units the dictator will transfer to the receiver in the role he has been assigned, i.e. will earn either the dictator's or the receiver's payoff. Using role uncertainty to implement the veil of ignorance removes the possibility to favor oneself over the other player and, at the same time, introduces risk. The *risk treatment* serves as a control treatment to isolate a subject's risk preferences. It has the same efficiency loss and role uncertainty as the veil of ignorance treatment, but it is a one person game. In the risk treatment each participant decides how to allocate the pie across the states of being dictator or being receiver and is randomly assigned the position of either dictator or receiver afterwards. However, the position not assigned to the decision maker is not filled in by a second person. The money assigned to the empty position is not paid out. The efficiency loss enables us to tell apart subjects with different degrees of risk aversion. In terms of risk, the decision situation in the risk and the veil of ignorance treatment is identical, but impartial social preferences can only be an additional motive in the two person veil of ignorance treatment.

By comparing decisions in the risk and the veil of ignorance treatment we can test our hypothesis that the difference principle can be derived from any degree of risk aversion as long as impartial social preferences for equality are sufficiently strong. If decisions in these two treatments don't differ significantly only risk aversion determines behavior behind the veil of ignorance. Hence, the claim that the difference principle can only be derived from infinite risk aversion is correct. If, in contrast, differences between the two treatments are significant and impartial social preferences in the veil of ignorance treatment reflect equality concerns, then the difference principle is com-

[6] In this paper, we define a more efficient allocation to be an allocation with a higher sum of payoffs of both players (Kaldor-Hicks efficiency).

patible with any degree of risk aversion if impartial social preferences for equality are sufficiently strong.

We find that subjects transfer significantly more in the veil of ignorance than in the dictator game treatment. Still, in the veil of ignorance treatment only a minority of subjects opts for the difference principle. In all three treatments we observe striking gender differences: women are more risk averse and have a stronger concern for equality than men. For men behavior does not differ significantly in the risk and the veil of ignorance treatment, i.e. for the vast majority of male subjects the veil of ignorance introduces only risk. In contrast, for women, impartial social preferences for equality are a second significant motivation besides risk in the veil of ignorance treatment. Our results for women imply that the difference principle can also be derived from impartial social preferences for equality and thus does not necessarily imply infinite risk aversion.

Some other economic experiments implement the veil of ignorance. Johannesson and Gerdtham (1995), Beckman et al. (2002), Johansson-Stenman, Carlsson and Daruvala (2002), and Carlsson, Gupta and Johansson-Stenman (2003) basically let subjects who do not yet know the place they (or their imaginary grandchildren) will occupy in a given society choose between societies that differ with respect to mean and distribution of income. Ackert, Martinez-Vazquez and Rider (2004) ask subjects to vote in favor of either a lump-sum or a progressive tax regime before they are randomly assigned a pre-tax payoff. To be able to interpret the observed behavior in terms of impartial social preferences, all mentioned experiments have to assume that subjects are risk neutral. Otherwise, the observed behavior can only be interpreted as the result of either risk aversion *or* impartial social preferences. The new contribution of our experiment is that we are able to separate the effects of risk aversion and impartial social preferences in a veil of ignorance setting.[7]

Only few further experiments in economics have elicited impartial social preferences without referring to the veil of ignorance. In Engelmann and Strobel (2004) one of the decision maker's tasks is to choose among three different allocations of payoffs across

[7] The veil of ignorance has also been the subject of experimental inquiries in other disciplines such that political sciences and psychology (Brickman, 1977; Curtis, 1979; Frohlich, Oppenheimer and Eavey, 1987; Bond and Park, 1991; Mitchell et al., 1993).

himself and two further subjects that represent an efficiency-equality trade off. Since the decision maker's payoff is constant across all three allocations, the experimental design controls for self-interest. A constant payoff for the decision maker also implies that his choice has no monetary consequences for himself. In contrast, one crucial aspect of the veil of ignorance, our object of investigation, is that both the decision maker and his reference group are affected by choices made behind the veil of ignorance.[8]

The remainder of the chapter is organized as follows. The details of the experimental design and implementation are explained in section 2. Section 3 presents the hypotheses to be tested and links them to the experimental design. Results are provided in section 4 that also elaborates on the striking differences in the behavior of male and female subjects. In the last section, we conclude. The appendix contains instructions, control questions and the experimental data.

1.2 Experimental Design and Procedure

1.2.1 The three treatments

The experimental design is based on a dictator game. Since the receiver is purely passive, the dictator game is one of the simplest ways to elicit the dictator's social preferences that do not interfere with any strategic considerations. In our experiment, dictators have to decide how to split a 12 unit pie.

We use a three treatment design. The *dictator game treatment* is a standard dictator game with one additional feature, an efficiency loss of 50 % for units transferred from the dictator to the receiver. The efficiency loss introduces a trade-off between equality and efficiency and can be interpreted as a deadweight loss that arises as the cost of redistribution. We choose an efficiency loss of 50 % because it is easy to calculate for the experimental subjects and makes our results comparable to those obtained in

[8] Being affected by one's own choice might influence behavior: First, the decision maker has monetary incentives to reveal his true preferences. Second, imagine a decision maker who prefers a very efficient, but highly unequal allocation. In a setup with a constant payoff for the decision maker, choosing the unequal allocation corresponds to "punishing" some of the other subjects while being on the safe side himself. In contrast, in our experiment, the decision maker himself risks getting a very low payoff when choosing an unequal allocation.

Andreoni and Miller (2002) and Andreoni and Vesterlund (2001). Since the dictator can only transfer integer units, the following allocations are possible results of the game:

Table 1.1: Possible allocations

dictator	12	11	10	9	8	7	6	5	4	3	2	1	0
receiver	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6

There are two focal points among these allocations: the allocation (12,0) represents the most efficient one (and, at the same time, the one a selfish dictator would choose). An individual with a very strong concern for equality would choose allocation (4,4). Transferring more than necessary to achieve the equal split allocation (4,4) is hard to rationalize: the resulting allocations impose an enormous efficiency loss and add inequality. The dictator game treatment serves as benchmark, ensures comparability with related studies and measures social preferences.

The *veil of ignorance treatment* implements the veil of ignorance by introducing role uncertainty. It is a dictator game with the same 50 % efficiency loss as the dictator game treatment and additional role uncertainty. Role uncertainty means that first every subject decides how many units the dictator will transfer. *After this transfer decision roles (dictator and receiver) are randomly assigned* and pairs consisting of one dictator and one receiver matched. Finally, a subject that has been assigned the receiver (dictator) role will be paid the receiver's (dictator's) payoff according to his *own* decision how many units the dictator will transfer to the receiver. For example, imagine a subject that has first decided that the dictator will transfer 4 units. If this subject then gets assigned the receiver role he will receive a payoff of $\frac{1}{2} \times 4 = 2$, his matched subject in the dictator role will receive $12 - 4 = 8$ units. If this subjects gets assigned the dictator role he will receive a payoff of 8 units, his matched subject in the receiver role will receive 2 units. It is possible that *every* subject's decision is implemented as the dictator's choice (independent of whether the decision-maker has been assigned the role of dictator or receiver) because each subject also serves as a dummy player in another subject's decision. Procedural details of our matching protocol are provided below.

Implementing the veil of ignorance as described above induces risk and potentially impartial social preferences. To test whether the veil of ignorance is only a concept about risk we have to be able to isolate potential impartial social preferences from risk considerations that jointly determine subjects' decisions behind the veil of ignorance.

The *risk treatment* serves exactly this purpose. It differs from the veil of ignorance treatment in just one respect. It is a one person game and consequently, basically a lottery decision: first, each subject decides how to allocate the pie across the states of being dictator or being receiver. After that decision each subject is randomly assigned the role of dictator or receiver with equal probability. In contrast to the veil of ignorance treatment, there is no second subject who fills in the role that has not been assigned to the decision-maker. For example, imagine a subject that has first decided that the dictator will transfer 4 units. If this subject then gets assigned the receiver role it will receive a payoff of $\frac{1}{2} \times 4 = 2$, $12 - 4 = 8$ units will not be paid out. If this subjects gets assigned the dictator role it will receive a payoff of 8 units, 4 units will not be paid out. Since there is no second subject who is affected by the decision maker's choice, the decisions in the risk treatment simply reflect the individual degree of risk aversion and cannot be influenced by social preferences.

Table 1.2 summarizes the three treatments.

Table 1.2: The three treatment design

treatment	characteristics			what is measured?
	efficiency loss	role uncertainty	number of players	
dictator game	yes	no	2	social preferences
veil of ignorance	yes	yes	2	impartial social preferences with risk
risk	yes	yes	1	risk attitude

There are two reasons that make the efficiency loss an essential feature of our experimental design: First, in the risk treatment the efficiency loss introduces a cost of insurance which allows telling apart risk neutral and risk averse subjects as well as risk averse subjects with different degrees of risk aversion. With any efficiency loss, risk neutral subjects who maximize their expected payoff strictly prefer the (12,0) allocation over all other allocations. For each risk averse subject, we get an approximate measure of individual risk aversion: the more risk averse a subject is the more units $0 < x < 8$ he will transfer. Very strongly risk averse subjects transfer 8 units which results in the (4,4) allocation that provides full insurance. Without the efficiency loss, all possible allocations would have an expected payoff of 6 and every risk averse subject would choose the (6,6) allocation that provides full insurance at no cost. Risk neutral subjects would be indifferent between all possible allocations and thus might also choose the (6,6) allocation.

Second, to test whether the difference principle can also be derived from impartial social preferences for equal outcomes we have to be able to observe whether less than infinitely risk-averse subjects (i.e. subjects who transfer $x < 8$ in the risk treatment) transfer $x = 8$ in the veil of ignorance treatment. Since the only difference between the veil of ignorance and the risk treatment is the existence of the second person a higher transfer in the veil of ignorance treatment is caused by a concern for equality. With a 50 % efficiency loss, only few subjects will opt for full, but very costly insurance in the risk treatment. For all but these very strongly risk averse subjects there is still room for moving towards a more equal allocation in the veil of ignorance treatment. In contrast, if there was no efficiency loss, all risk averse subjects would choose the (6,6) allocation in the risk and the veil of ignorance treatment irrespective of whether they are purely selfish or have impartial social preferences for equality. Only for those risk neutral subjects who would transfer $x < 6$ in the risk treatment we could learn whether they have impartial social preferences for equality behind the veil of ignorance.[9]

1.2.2 Sessions

Due to matching requirements each subject participated in two of the three treatments: in the risk treatment and in one of the two two-player treatments, either the dictator game or the veil of ignorance treatment. At each time of the experiment half of the

[9] The experimental design cannot distinguish between subjects who are risk neutral and those who are risk loving. Both will choose the (12,0) allocation. This might be a flaw as, ceteris paribus, a more risk loving individual will let the dictator transfer less in the veil of ignorance treatment, a decision that we will interpret to reflect a preference for efficiency. We do not expect many subjects to be risk loving.

subjects played the risk treatment. These subjects were matched with the other half of subjects who played one of the two two-player treatments in the same room at the same time. This matching across treatments has two advantages: first, not only in the risk but also in the veil of ignorance and the dictator game treatment *every* subject's decision is in fact implemented (and every subject knows this[10]). We avoid introducing an additional source of risk in the veil of ignorance treatment, namely whether one's own decision or the decision of one's matched subject will be implemented. Second, we maximize the number of observations because we avoid paying passive players. As a result of the matching, each subject had three sources of payoff at the end of the session: the payoff from his own risk decision, a payoff from his own decision in one of the two two-player treatments and a payoff from a randomly assigned subject's decision in one of the two two-player treatments. Subjects were only informed about the last, additional source of payoff that they could not influence anyway at the end of the experiment.

In total we conducted nine sessions. In five sessions, all subjects played the risk and the veil of ignorance treatment, though in different orders. In the remaining four sessions, half of the participants first played the risk and then the veil of ignorance treatment, while the other half of participants first played the dictator game and then the risk treatment. The three treatment orders are depicted in table 1.3.

Table 1.3: Treatment orders

first treatment	second treatment	number of subjects
risk	veil of ignorance	83
veil of ignorance	risk	48
dictator game	risk	36

Before we pool the data obtained in one specific treatment, but from different treatment orders we have to make sure that there are no order effects. For the two treatment orders of the veil of ignorance treatment we use the Mann–Whitney test and for the three treatment orders of the risk treatment we use the Kruskal-Wallis test

[10] We told decision-makers in the dictator game and the veil of ignorance treatment in the instructions: "only half of the participants present in this room are taking part in the same experiment as you do. The other half of the participants is playing another experiment whose payoff does not affect you at all. You are assigned a participant from this other half".

to check whether the distributions of transferred units obtained in different treatment orders are significantly different. Table 1.4 shows that we can pool all veil of ignorance and risk treatment data respectively for the whole sample as well as for men and women separately.[11] In sum, we collected 131 observations on decisions in the veil of ignorance treatment, 167 in the risk and 36 in the dictator game treatment. The complete experimental data are displayed by treatment and sex in Appendix 1.6.2.

Table 1.4: No order effects

	veil of ignorance treatment (Mann-Whitney test*)	risk treatment (Kruskal-Wallis test*)
all	p=0.627 (131 obs.)	p=0.464 (167 obs.)
men	p=0.505 (91 obs.)	p=0.816 (108 obs.)
women	p=0.810 (40 obs.)	p=0.729 (59 obs.)

*: The reported p-values refer to two-tailed tests and are adjusted for ties.

1.2.3 Experimental procedure and subjects

The order of events during each experimental session was the following: Subjects were welcomed and randomly assigned a cubicle in the laboratory where they took their decisions in complete anonymity from the other subjects. The random allocation to a cubicle also determined the individual treatment order. Subjects were handed out the instructions for their first treatment and answered several computerized control questions that tested their understanding of the decision situation. Only after providing and explaining the right answers on the computer screen, we proceeded to the decision stage of the first treatment. After all subjects had made their first decision, we announced that there would be a second and at the same time last experiment. To avoid income effects we did not give subjects any feedback on the result of the first treatment before they were paid at the end of the whole session. The second treatment followed with the same procedures. We finished each experimental session by asking subjects to answer a questionnaire on their demographic characteristics, the strategies they had used and their expectations concerning the behavior and risk attitudes of the

[11] We present test results also by sex as gender differences will be important for the interpretation of our results.

other subjects.

A translated version of the instructions and the corresponding control questions can be found in Appendix 1.6.1. The experiment was programmed using the experimental software zTree (Fischbacher, 1999) and conducted at the experimental laboratory of the SFB 504 at the University of Mannheim, Germany in November 2005. The experiments lasted about one hour and subjects earned about 16 Euros on average. All 167 subjects[12] were university students with a large variety of subjects. The main characteristics of the subjects are displayed in Table 1.5.

Table 1.5: Composition of treatments

	dictator game treatment	risk treatment	veil of ignorance treatment
number of observations	36	167	131
sex*	19 (F)/17(M)	59(F)/108(M)	40(F)/91(M)
mean age	23.56	23.77	23.82
knowledge in economics**	66.67 %	64.67 %	64.12 %

* F stands for female, M for male
** includes students studying economics or business administration as minor or major

1.3 Hypotheses

Let us first briefly turn to the growing literature on gender differences in risk attitudes and social preferences. Reviewing the vast economic literature on gender differences in risk preferences Eckel and Grossman (2006) conclude that women are characterized by a higher degree of risk aversion than men in field studies, while the results from laboratory experiments are less consistent. Similarly, Croson and Gneezy's (2004) survey summarizes that there is clear evidence that men are more risk-taking than women in most tasks and most populations. Camerer (2003, p.64) summarizes evidence on the effect of gender on social preferences and concludes that evidence is mixed.[13]

[12] We admitted only an even number of subjects to the experiment but one subject left during the course of the experiment. His role was filled by one of the experimenters and the corresponding observations were deleted.

[13] Considering dictator games, for example, Eckel and Grossman (1998) find that women on average donate twice as much as men in a standard dictator game. Similarly, Dufwenberg and Muren (2005)

However the studies that are most closely related to our dictator game and veil of ignorance treatment indicate that gender differences are likely to matter in our experimental setup. In Andreoni and Vesterlund (2001) subjects play dictator games with different levels of efficiency losses. They find that when it is relatively expensive to give, women are more generous than men. As the price of giving decreases, men begin to give more than women. With our 50 % efficiency loss, women are significantly more generous than men. The following two studies are, to some extent, related to our veil of ignorance treatment: they have an impartial decision maker as we do, but, in contrast to our study, the decision maker's payoff is fixed and independent of his own choice. Fehr, Naef and Schmidt (2006) replicate Engelmann and Strobel's (2004) experiment and find that women choose the most egalitarian allocation significantly more often than men. Dickinson and Tiefenthaler (2002) play an experiment with a disinterested third-party decision maker in which women are significantly more likely to choose an allocation resulting in equal payoffs while men are more likely to choose the most efficient allocation.

To check for the existence of gender differences in our experimental setup we formulate Hypothesis 1:

Hypothesis 1

Women and men do not behave significantly different in any of our treatments.

If we should reject hypothesis 1 gender differences in risk attitudes and social preferences are likely to affect all further results on differences between treatments. Consequently, we should then analyze the following hypotheses not only for both sexes jointly, but also for men and women separately.

Exploiting our three treatment design we can first compare transfers in the dictator game and the veil of ignorance treatment that have the same trade off between equality and efficiency.

present results of a dictator game in which significantly fewer men than women give non-zero amounts. In contrast, Bolton and Katok (1995) find no systematic gender differences in a standard dictator game.

Hypothesis 2

There is no significant difference between social preferences and impartial social preferences with risk that are measured in the dictator game and the veil of ignorance treatment respectively.

If we should reject hypothesis 2, we will ask next whether the observed difference can be completely explained by risk aversion: Is the veil of ignorance only a concept that introduces risk? Or, in contrast, are impartial social preferences an additional motivation behind the veil of ignorance?

Hypothesis 3

There is no significant difference between risk preferences and impartial social preferences with risk that are measured in the risk and the veil of ignorance treatment respectively.

The only difference between the risk and the veil of ignorance treatment is whether a second person exists who is affected by the decision maker's transfer decision. Since the degree of risk is held constant, the two treatments differ only in whether impartial social preferences can possibly motivate the observed transfer decisions. If we cannot reject hypothesis 3, the thought experiment of a veil of ignorance has correctly been perceived as a concept inducing only risk aversion. The only way to derive Rawls' difference principle is to assume infinite risk aversion. In contrast, if hypothesis 3 is rejected, impartial social preferences are a significant motivation behind the veil of ignorance. Consequently, the difference principle and maximin preferences can also be considered the result of impartial social preferences combined with any degree of risk aversion (assuming that impartial social preferences induce an increased concern for equality).[14]

[14] While the term "veil of ignorance" was coined by Rawls, Harsanyi (1953, 1955) already used the same thought experiment. Harsanyi interprets value judgments made behind the veil of ignorance to reflect choices involving just risk and assumes that agents are risk neutral. Consequently, he predicts efficiency seeking behavior to prevail behind the veil of ignorance. In terms of our experiment, Harsanyi's argument would be supported if we found that subjects do not transfer any units in the risk treatment (risk neutrality) and if differences in subjects' behavior across the risk and the impartiality treatment were not significant.

This is investigated by hypothesis 4: given that impartial social preferences introduce an additional motive, do they induce an increased concern for equality or for efficiency? To what extent does a veil of ignorance like situation induce maximin preferences as predicted by Rawls?

Hypothesis 4

In the veil of ignorance treatment subjects behave according to maximin preferences.

1.4 Results

1.4.1 Gender differences

Result 1

In our experiment, women are significantly more risk averse than men and choose more equal (and thus less efficient) allocations than men.

In total, we had 108 male (65 %) and 59 female (35 %) subjects. Table 1.6 displays average transferred units by sex and treatment as well as test results by treatment for whether medians and distributions of transferred units differ for men and women.

Table 1.6: Gender differences by treatment

treatment	mean men	mean women	Mann-Whitney test*	Median test*
dictator game	0.76 (17 obs.)	2.37 (19 obs.)	p=0.061	p=0.091**
risk	2.72 (108 obs.)	3.69 (59 obs.)	p=0.014	p=0.016
veil of ignorance	2.81 (91 obs.)	5.00 (40 obs.)	p=0.000	p=0.000

*: The reported p-values refer to two-tailed tests and are adjusted for ties.
**: In the dictator game treatment, the median corresponds to keeping all 12 units. To obtain a test result we treat observations that equal the median like observations greater instead of lower than the median as we do in all other Median tests reported.

In sum, we observe strikingly different transfer behaviors of male and female subjects: according to Mann-Whitney tests the distributions of units transferred differ significantly for men and women both in the risk and in the veil of ignorance treatment. The same is true for medians. In the veil of ignorance treatment, the absolute difference in means is largest and amounts to 2.2 units with 8 units being the maximal reasonable transfer amount. Women transfer more and thus are more concerned about equality while men care more about efficiency. Gender differences in the risk treatment are smaller in absolute amounts, but strongly significant: they indicate that, on average, women are more risk averse than men. Due to the small number of observations medians and distributions are only weakly marginally different in the dictator game treatment. Still, on average male dictators transfer less than one out of 8 units, female dictators transfer nearly 2.5 units. Furthermore, about 70 % of male dictators keep the whole pie, while only 37 % of women do. Carlsson, Daruvala and Johansson-Stenman (2005) run two different treatments to measure a given individual's risk and inequality aversion in the absence of risk. Similar to our results, they find that female subjects are more risk averse and more inequality averse than men.

In sum, male and female subjects do behave significantly different in our experiment. Consequently, we will focus on analyzing the data for men and women separately. We will also present a joint analysis for the sake of completeness and to guarantee comparability of our results in the dictator game treatment to other dictator game studies.[15]

1.4.2 Comparison of dictator game and veil of ignorance treatment

We now turn to hypothesis 2 and discuss whether stated preferences in front of and behind the veil of ignorance differ. If they do, we might want to question the use of people's stated social preferences from surveys and alike as a basis for "just" policy

[15] Our results in the dictator game treatment are very close to those of other dictator games that vary the price of giving. In our dictator game treatment, subjects give away 13 % of the pie on average. With the same 50 % efficiency loss and a similar pie size, they transfer 10 % in Andreoni and Vesterlund (2001) and 21 % in Andreoni and Miller (2002). In Fisman, Kariv and Markovits (2007), for an efficiency loss of 30 % or above, 60 % of subjects transfer less than 5 % of the pie, 17 % transfer 5-15 % of the pie, 10 % 15-25 % of the pie and the remaining subjects transfer more. The corresponding figures in our dictator game treatment are very similar: 53 %, 17 % and 11 %, respectively.

design. Our data would then suggest using impartially stated social preferences.

Result 2

There is a large and significant difference between social preferences and impartial social preferences with risk. Subjects transfer significantly more in the veil of ignorance than in the dictator game treatment.

Table 1.7: Test results for hypothesis 2

	mean dictator game treatment	mean veil of ignorance treatment	Mann-Whitney test*	Median test*
all	1.61 (36 obs.)	3.48 (131 obs.)	p=0.000	p=0.000
men	0.76 (17 obs.)	2.81 (91 obs.)	p=0.003	p=0.018
women	2.37 (19 obs.)	5.00 (40 obs.)	p=0.002	p=0.000

*: The reported p-values refer to two-tailed tests and are adjusted for ties.

Test results in Table 1.7 reject hypothesis 2: medians and distributions of units transferred differ significantly for the pooled data and for men and women separately. OLS regression results using the pooled dictator game and veil of ignorance treatment data in the first two columns of Table 1.8 confirm the test results: both men and women transfer significantly more in the veil of ignorance than in the dictator game treatment, about 2 units on average.[16] In both treatments, women transfer significantly more than men, a bit but not significantly more so in the veil of ignorance treatment.

One would have expected hypothesis 2 to be true only if (i) experimental subjects were risk neutral and (ii) they would behave impartially even if their role is known, i.e. if experimental subjects would not exhibit any egoism or subconscious self-serving bias in the dictator game treatment. Thus, the next step is to figure out where the significant differences between the dictator game treatment and the veil of ignorance treatment stem from: Are they due to risk aversion only, the prevalence of impartial social preferences in the veil of ignorance treatment as opposed to egoism in the dictator

[16] Curtis (1979) also compares individual distributional preferences in front of and behind the veil of ignorance but adds the issue of meritocracy: subjects have to decide how to distribute 3 dollars between a high and a low scorer in a motor skill test. When subjects know whether they are the high or the low scorer, 13 % behave consistently with maximin preferences, when they do not know 52 % do. Hence, concerns for equality are also stronger behind than in front of the veil of ignorance.

Table 1.8: Pooled OLS

dependent variable: transfer amount explanatory variables*	dictator game and veil of ignorance treatment data		risk and veil of ignorance treatment data	
	coefficient	p-value**	coefficient	p-value**
female	1.606	0.054	1.091	0.018
VoI	2.064	0.000	-0.059	0.901
female x VoI	0.741	0.443	1.308	0.053
sequence risk - VoI	-0.121	0.826	0.616	0.911
sequence VoI - risk	-	-	-0.166	0.802
VoI x sequence VoI** - risk	-	-	0.343	0.651
economist	0.242	0.599	0.397	0.262
age (in years)	0.028	0.961	-0.198	0.708
age squared	0.001	0.949	0.005	0.609
constant	-0.439	0.953	0.005	0.609
N	167		298	
R^2	0.176		0.087	

*: female = 1 if female, 0 if male; VoI = 1 if veil of ignorance treatment, 0 else;
risk = 1 if risk treatment, 0 else; economist = 1 if economist, 0 else
**: based on robust standard errors

game treatment, or a combination of both? In the risk treatment, 68 % of all subjects (80 % of female and 61 % of male subjects) transfer a positive amount despite the large efficiency loss occurred. The average transfer amount is 3.1 for all subjects, 3.7 for women and 2.7 for men. The majority of our subjects clearly are risk averse.

1.4.3 Comparison of risk and veil of ignorance treatment

Can risk aversion account for the complete observed difference in transfers between the dictator game and the veil of ignorance treatment? Or do impartial social preferences additionally contribute to it?

Result 3

For female subjects impartial social preferences are a second significant motivation behind the veil of ignorance besides risk, while this is not true for men.

Table 1.9 compares all observations obtained in the risk and the veil of ignorance treatment. Analyzing only the data that are pooled for both sexes, we would conclude that hypothesis 3 cannot be rejected: both medians and distributions of transfer

Table 1.9: Test results for hypothesis 3

	mean risk treatment	mean veil of ignorance treatment	Mann-Whitney test*	Median test*
all	3.07 (167 obs.)	3.48 (131 obs.)	p=0.203	p=0.484
men	2.72 (108 obs.)	2.81 (91 obs.)	p=0.773	p=0.980
women	3.69 (59 obs.)	5.00 (40 obs.)	p=0.011	p=0.047

*: The reported p-values refer to two-tailed tests and are adjusted for ties.

amounts do not differ significantly across the two treatments. The veil of ignorance treatment dummy is not significant in the right part of Table 1.8 that presents OLS regression results for pooling all risk and veil of ignorance treatment data. However, taking a closer look at the data we find that there are striking gender differences. While hypothesis 3 cannot be rejected for men at all, it actually can be rejected for women. For the female subjects, medians and distributions of transfer amounts do differ significantly in the risk and the veil of ignorance treatment. The regression results in Table 1.8 document that women transfer significantly (1.3 units) more in the veil of ignorance treatment than in the risk treatment.

In sum, for female subjects impartial social preferences are a major motivation behind the veil of ignorance, while this is not true for men. Impartial social preferences seem to increase equality concerns.

To check Rawls' prediction that maximin preferences prevail behind the veil of ignorance, Table 1.10 categorizes the data according to "strong types", i.e. the share of subjects who decide in favor of full efficiency or full equality in each of the two treatments.

Table 1.10: Strong types

	risk treatment			veil of ignorance treatment		
participants choosing		percentage	number		percentage	number
full efficiency	all	32.3 %	54/167	all	27.5 %	36/131
	men	38.9 %	42/108	men	35.2 %	32/91
	women	20.3 %	12/59	women	10.0 %	4/40
full equality, full insurance	all	4.2 %	7/167	all	13.7 %	18/131
	men	3.7 %	4/108	men	8.8 %	8/91
	women	5.1 %	3/59	women	25.0 %	10/40

Result 4

In the veil of ignorance treatment, only 8.8 % of men and 25.0 % of women act according to maximin preferences. Still, for women impartial social preferences clearly induce a concern for equality.

We observe that nearly all subjects react to the large efficiency loss in the risk treatment: only very few subjects choose full insurance by equalizing payoffs across states. In the veil of ignorance treatment, the share of subjects choosing full equality increases substantially, it doubles for men and is five times as high for women. Still, support for Rawl's difference principle is only limited: 8.8 % of men and 25.0 % of women choose full equality of payoffs. Related experiments that elicit paid impartial decisions behind the veil of ignorance also find low support for maximin preferences. In Carlsson, Gupta and Johansson-Stenman (2003) and Johansson-Stenman, Carlsson and Daruvala (2002) only 20 % and 19 % of subjects act in a way that is compatible with the difference principle. In Frohlich, Oppenheimer and Eavey (1987), who investigate paid group decisions, no group ever chooses an income distribution that maximizes the lowest income. Maximizing the average income plus a floor constraint is the most popular principle for choosing among income distributions. In contrast, in Curtis (1979) 52 % of subjects behave according to maximin preferences behind the veil of ignorance, in Mitchell et al. (1993) with unpaid decisions and compulsory participation between 65 % and 83 % of subjects (for differing degrees of meritocracy) opt for the difference principle.

In our experiment, a bit more than one third of men go for full efficiency in both the risk and the veil of ignorance treatment. In sharp contrast, the share of women opting for full efficiency halves in the veil of ignorance treatment. Compared to the situation in the one-person risk treatment, full efficiency now implies maximal inequality. All these findings underline major differences in the behavior of men and women: they show that in our experiment, women exhibit impartial social preferences for equality in a much stronger way than men.

The results presented above are confirmed by a within subject analysis where we compare a given individual's decision in the risk and the veil of ignorance treatment (and thus skip risk treatment data from the dictator game treatment - risk treatment sequence). Applying a Wilcoxon signed rank test to the pooled within subject data (131 observations) yields p=0.037 (two-sided), i.e. distributions of transfer amounts differ significantly in the risk and the veil of ignorance treatment. This result is purely caused by the behavior of female subjects. A two-sided Wilcoxon signed rank test reveals that female subjects transfer significantly different amounts in the risk and veil of ignorance treatment (p=0.006) while men do not (p=0.790).

Table 1.11 classifies subjects according to three "weak types", namely whether an individual does not react at all to the existence of the second person in the veil of ignorance treatment, whether it opts for more equality or for more efficiency as soon as the second person shows up.

Table 1.11: Within subject analysis

subjects who transfer ...	all	men	women
the same amount in the risk and the veil of ignorance treatment	44 %	53 %	22.5 %
more in the veil of ignorance treatment	35 %	24 %	60 %
less in the veil of ignorance treatment	21 %	23 %	17.5 %
number of observations	131	91	40

For more than half of the male subjects the existence of the second person does not add impartial social preferences as a motive, while this is only true for less than 1/4 of female subjects.[17] For those male subjects for whom impartial social preferences matter their effect is equally likely to point in the direction of an increased efficiency or an equality motive. 60 % of women transfer more in the veil of ignorance treatment than in the risk treatment (3.1 units on average), but only about 1/4 of men do (4.0 units on average). These findings confirm that for the vast majority of female subjects the veil of ignorance induces impartial social preferences for equality besides inducing risk. Our results for those 14 out of 131 subjects (7 men and 7 women) who do not

[17]Subjects who transfer the same amount in both treatments could also have a degree of risk aversion and impartial social preferences that imply the same transfer amount in the veil of ignorance treatment. While we cannot totally disapprove this possibility, we can be sure that these subjects' decisions are, on average, not driven by strong equality concerns: they transfer only 2.2 out of 8 reasonably possible units in the veil of ignorance treatment.

opt for full insurance in the risk treatment, but choose full equality in the veil of ignorance treatment imply that the difference principle can be derived from impartial social preferences for equality and does not require that subjects are infinitely risk averse. Impartial social preferences for equality are even a more prominent motive for choosing the maximin allocation in the veil of ignorance treatment. Only 3 subjects act according to maximin preferences in the veil of ignorance treatment because they are extremely risk averse, i.e. transfer 8 units in both the veil of ignorance and the risk treatment. These results contrast the utilitarians' claim that maximin preferences necessarily represent preferences with infinite risk aversion. We should keep in mind, however, that overall support for the difference principle is only limited.

While our results for women demonstrate that impartial social preferences for equality are one important motive behind the veil of ignorance there are also subjects - 23 % of men and 17.5 % of women - with impartial social preferences for efficiency.[18] Insofar our results are related to those of Engelmann and Strobel's (2004) taxation games that document that both concerns for efficiency and maximin preferences are important motives for impartial decision makers.

1.5 Conclusion

Rawls' claim that a truly just allocation of resources can only be based on impartial judgments made behind the veil of ignorance is as intuitively attractive as disputable: democratic institutions rest upon the assumption that competition of vested interests is able to balance interests appropriately. It was not the aim of this paper to comment on this. Our experimental results simply show that preferences stated in front of and behind the veil of ignorance differ significantly. Behind the veil of ignorance,

[18] In our data, subjects who transfer less in the veil of ignorance treatment than in the risk treatment are substantially more risk averse than those who transfer more. A possible explanation for why subjects transfer less could be that subjects maximize the sum of their own and the second person's expected utility but do not have any distributional concerns. Subjects would then give away less (more) in the veil of ignorance treatment if they perceive themselves as more (less) risk averse than the average participant. In the final questionnaire we asked our subjects to assess whether they had transferred more or less than the average participant in the risk treatment. We run an OLS regression to explain the difference in transferred units in the veil of ignorance and the risk treatment. Controlling for subject characteristics, the individual perception of own risk aversion compared to average risk aversion is not significant. Consequently, our data reflect distributional concerns.

subjects prefer more equal distributions, but only a minority of subjects acts according to maximin preferences. Consequently, support for Rawls' difference principle is far from being unanimous. On a technical level, we have presented an experimental design that separates the effects of risk and impartial social preferences behind the veil of ignorance. We have found that men prefer more equal distributions mostly for insurance purposes. In contrast, women's choice of more equal allocations is also due impartial social preferences that value equality per se. Most importantly, our results for those subjects who act according to maximin preferences in the veil of ignorance, but not in the risk treatment challenge the utilitarians' claim that behind the veil of ignorance maximin preferences necessarily represent preferences with infinite risk aversion.

Our results also contribute to the growing literature on gender differences in economic behavior. Gender effects in our data are strong. They imply that women are more risk averse than men. Furthermore, when there is a trade off between equality and efficiency women seem to have stronger preferences for equal allocations while men have stronger preferences for efficient allocations.

1.6 Appendix

1.6.1 Instructions and control questions

Both instructions and control questions were originally in German. The translated instructions and control questions presented below are those of the veil of ignorance treatment. The instructions and control questions for the dictator game and the risk treatment are structured and phrased in the same way with just one exception: to explain the risk treatment in the most natural and easiest possible way the instructions did not mention the state of being participant A (dictator) or B (receiver), but described the two possible states by throwing a dice and getting either an even or an odd number. The instructions of the dictator game and the risk treatment are available from the author upon request.

Instructions

General explanations concerning the experiment

Welcome to this economic experiment.

If you read the following instructions carefully, you will be able to earn an amount of money that depends on your own decisions. Therefore, it is very important that you read these explanations carefully. If you have any questions, please do not hesitate to ask us. Please raise your hand, and we will come to your seat.

During the experiment you are not allowed to talk to the other participants, to use cell phones or to start any programs on the computer. The neglect of these rules will lead to the exclusion from the experiment and all payments.

During the experiment we talk about points instead of Euros. Your total income will therefore be calculated in points first. At the end of the experiment, the total amount of points obtained during the experiment will be converted in Euros at an exchange rate of

1 point = 1 Euro.

At the end of the experiment, you will be paid your earned income that is the result of your decision in cash.

On the next pages we will explain the exact course of the experiment.

The Experiment

In this experiment there are **two participants**, A and B.

Participant A has an initial endowment of 12 points, whereas participant B has an initial endowment of 0 points. Participant A can transfer every integer amount between 0 and 12 points (0 and 12 included) to participant B. Every transfer leads to the loss of half of the transferred points. **This means that participant B receives only half of a point for every full point participant A transfers to him.** Participant

B does not have any influence on the decision of participant A and the course of the game apart from being paid half of the points transferred to him by participant A at the end of the experiment. Participant A will be paid the amount of points that he does not transfer.

The following table shows all possible distributions of points for participant A and B at the end of the experiment:

A transfers to B	0	1	2	3	4	5	6	7	8	9	10	11	12
A's points	12	11	10	9	8	7	6	5	4	3	2	1	0
B's points	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6

The course of the experiment is the following:

Stage 1:

First, you have to decide how many points participant A transfers to participant B. This can be done by entering the number of points that are transferred from participant A to participant B on the following screen and pushing the "OK"-Button afterwards. **Note that at this stage you do not know yet whether you will be a participant A or a participant B in stage 2.** The computer has already randomly chosen another participant with whom you form a pair.

[screen]

Stage 2:

A random selection determines whether you are assigned the role of participant A or the one of participant B. When you are assigned the role of participant A, the participant assigned to you has the role of participant B. When you are assigned the role of participant B, the participant assigned to you has the role of participant A. **Every pair therefore consists of one real participant A and one real participant B.** Both during the experiment and afterwards neither you nor the participant assigned to you know who the respective partner is.

Stage 3:

Your decision in stage 1 will be realized in any case, independent from whether you are assigned to the role of participant A or B. (This is possible because only half of the participants present in this room are taking part in the same experiment as you do. The other half of the participants is playing another experiment whose payoff does not affect you at all. You are assigned a participant from this other half.)

Example 1: You decide that A transfers 5 points to B. B therefore obtains 5:2=2.5 points and A keeps 12-5=7 points. Afterwards, it is decided by drawing lots that you are participant B. Your decision is implemented: You obtain 2.5 points. The participant assigned to you obtains 7 points.

Example 2: You decide that A transfers 5 points to B. B therefore obtains 5:2=2.5 points and A keeps 12-5=7 points. Afterwards, it is decided by drawing lots that you are participant A. Your decision is implemented: You obtain 7 points. The participant assigned to you obtains 2.5 points.

This experiment is played only once. At the end of the experiment all participants A and B are paid their income **in cash**.

If you have any questions, please raise your hand. We will come to your seat to answer your question.

Control questions

Question 1: You decide that A transfers 3 points to B. It is decided by drawing lots that you are participant A.

How many points does B get?

How many Euros will you be paid?

How many Euros will your randomly assigned participant B be paid?

Question 2: You decide that A transfers 6 points to B. It is decided by drawing lots that you are participant B.

How many points does B get?

How many Euros will you be paid?

How many Euros will your randomly assigned participant A be paid?

1.6.2 Data by treatment and sex

Figure 1.1: Transferred amount in the dictator game treatment by sex

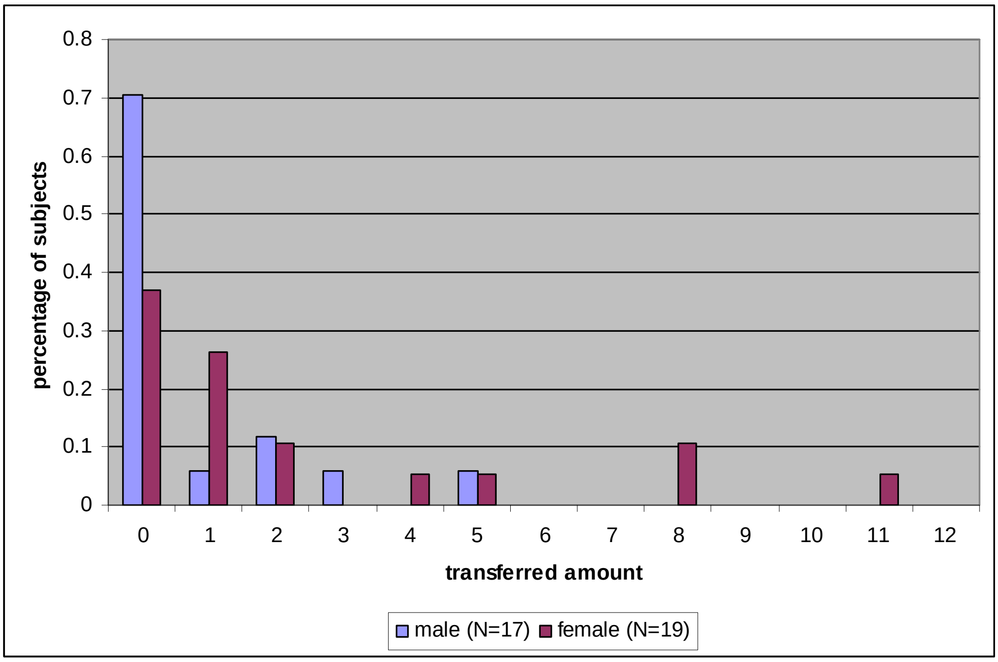


Figure 1.2: Transferred amount in the risk treatment by sex

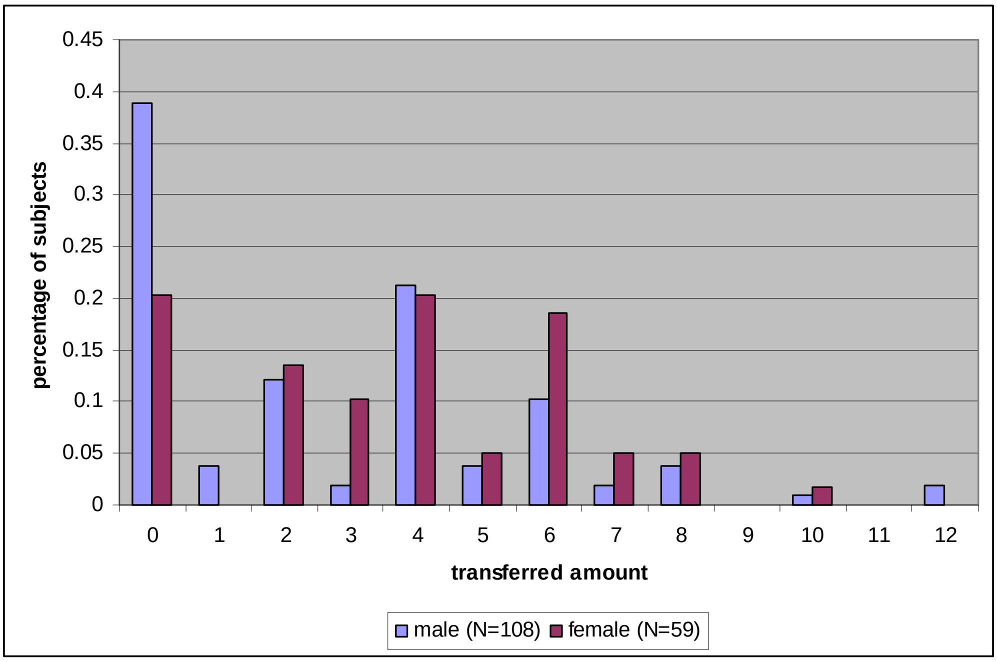
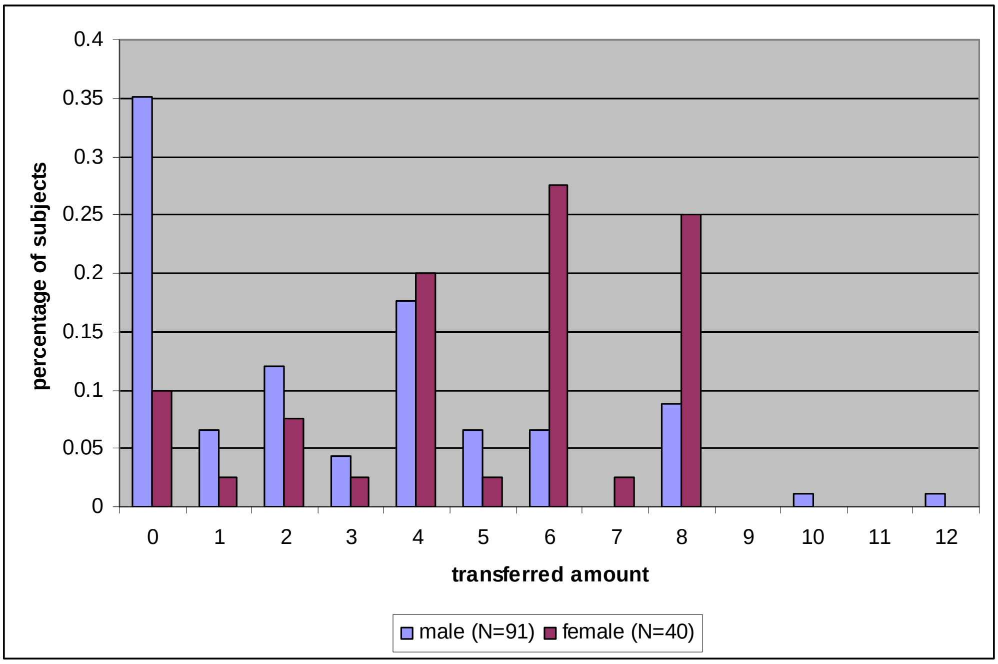


Figure 1.3: Transferred amount in the veil of ignorance treatment by sex



Chapter 2

An experimental test of the deterrence hypothesis*

2.1 Introduction

That crime has to be punished seems to be universally accepted. The purpose and level of punishment, however, are controversial. Immanuel Kant advocated punishment to re-establish justice, Georg Friedrich Wilhelm Hegel stressed that ill has to be retaliated with ill. Both philosophers regard punishment as a mean to establish justice. Becker's (1968) deterrence hypothesis is the classic economic contribution to the debate on punishment. According to Becker the purpose of punishment is to (efficiently) deter individuals from committing crimes. To achieve deterrence Becker relies on the power of pure deterrent incentives such as the severity and probability of punishment. The deterrence hypothesis states that crime rates fall in the severity and in the probability of punishment.

Our laboratory experiment directly tests the deterrence hypothesis in a controlled environment that permits to exogenously vary deterrent incentives, i.e. detection probability and level of punishment. For this purpose we use a straightforward context, namely subjects have the possibility to steal from another subject's payoff. They cannot only decide whether to steal or not to steal, but also how much to steal. We ask a

*This chapter is joint work with Christina Strassmair from the University of Munich.

very basic, but important question: do deterrent incentives work?

In order to answer this question we have chosen one of the simplest possible designs: a modified dictator game. Two agents, A and B, are randomly matched. Agent A is a passive agent and has a higher initial endowment than agent B. Agent B can decide how much to take away (steal) from A's initial endowment. With probability $1-p$, this amount is transferred from A to B. With probability p ("detection probability"), however, this amount is not transferred and a fixed fine f is deducted from B's initial endowment if B has chosen a strictly positive transfer amount.

We conduct six different treatments in which we vary detection probability p and fine f. Our benchmark treatment T1 sets $p = f = 0$. Treatments T2, T3 and T4 investigate the range of small and intermediate deterrent incentives, i.e. levels of incentives such that taking agent A's whole initial endowment pays off in expectation. Treatment T5 is characterized by a combination of p and f such that taking everything generates about the same expected payoff as taking nothing. In treatment T6, however, the expected payoff from taking everything is substantially smaller than the one from taking nothing. Each subject participates in two different treatments sequentially. This design permits both an across subjects and a within subject analysis of taking behavior. In other words, we can analyze both different regimes and regime changes with the data at hand.

Our very simple experimental design has three main advantages. For subjects, the task is easy to understand. Our design allows testing the isolated effect of incentives. And our design captures some crucial features of many crimes: the victim is rather passive. It cannot set the severity of punishment and - to a large extent - the detection probability. In case of theft, the stolen amount is a good predictor of the thief's benefit and the victim's cost.

The results of our across subjects analysis clearly reject the deterrence hypothesis: the average taken amount is not monotonically (weakly) decreasing in p and f. In contrast, we find that incentives may backfire: on average subjects take significantly more in treatments with intermediate deterrent incentives than in the absence of incentives. Only very strong incentives deter. Both our across and within subjects results can be explained by a model of two types: selfish subjects who react to deterrent incentives as predicted by the deterrence hypothesis and fair-minded subjects who take more when incentives are introduced or raised until incentives reach a very high level. Possible explanations for the behavior of fair-minded subjects are crowding out of fairness concerns by extrinsic incentives or fairness preferences with respect to expected outcomes. Only lasting crowding out of fairness concerns can explain the order effects in our data: many fair subjects take more in a given treatment if this treatment was preceded by a treatment with stronger incentives than if it was preceded by a treatment with weaker incentives. Furthermore, we find that p and f seem to be interchangeable instruments in achieving deterrence.

Since we obtain our data from neutrally framed experiments (i.e. we talk about "transfer decisions" instead of "stealing" and about "minus points" instead of a "fine"), one may question the applicability of our results for "real life crime". In real life, crime and deterrent incentives often have a strong moral connotation, and policy makers may make use of that. Still, we consciously use a neutral frame because our primary aim is to test the economic approach to crime. Its core, the deterrence hypothesis, relies on pure incentive effects that are independent of any frame. In Becker's (1968) model framing might ceteris paribus affect B's decision, but not the comparative statics with respect to p and f. Whatever the frame the taken amount should be monotonically decreasing in p and f. To measure the effect of moral costs we run some additional "morally" framed sessions in which we label transfers as stealing and talk about a fine instead of minus points. In these sessions, we still observe backfiring of incentives.

Becker's seminal paper has triggered numerous theoretical extensions as well as field studies testing its external validity.[19] At large the empirical literature implies that punishment reduces crime, but variations in detection probability and severity of punishment explain only a small part of the variation in crime (see Glaeser, 1999). This may be caused by methodic problems that arise when using field data. Usually only aggregate data are available which results in simultaneity bias[20] and omitted variable

[19] Garoupa (1997) and Polinsky and Shavell (2000a) provide comprehensive overviews on the economic theory of optimal law enforcement. Eide (2000) and Glaeser (1999) survey empirical studies of the deterrence hypothesis.

[20] See Levitt (1997) for a convincing example of how to address the simultaneity problem.

problems. Field data often report the behavior of offenders only and not that of the general population. Furthermore, measurement error is widespread as not all crime is reported. All these problems do not exist in the laboratory.

Our experiment is related to the experimental literature on tax evasion that explicitly addresses deterrence. The tax evasion setups clearly differ from ours though: in many settings subjects do not influence other subjects' payoffs at all, in other settings collected taxes are used for public good provision or redistribution of resources among a group of subjects. In contrast, in our setup a stealing subject directly hurts another subject which seems to be a crucial feature of many crimes.[21] Laboratory experiments on criminal behavior other than tax evasion are scarce. While we explicitly test whether deterrent incentives work, other studies on criminal behavior simply assume that deterrent incentives work and focus on more elaborate issues. Falk and Fischbacher (2002) explore the influence of social interaction phenomena on committing a crime. Bohnet and Cooter (2005), Tyran and Feld (2006), and Galbiati and Vertrova (2005) investigate whether law can act as "expressive law", i.e. prevent crime by activating norms that prohibit committing a crime. Tyran and Feld (2006) also compare the effects of exogenously imposed and endogenously chosen incentives in a public good setting.

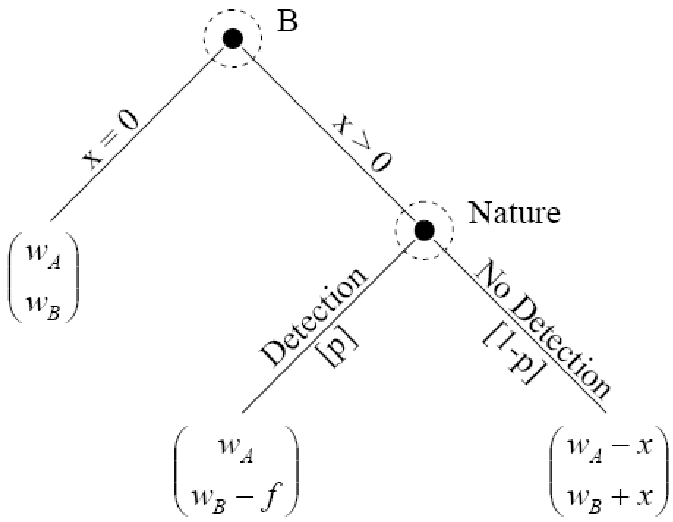
In addition, there is a growing economic literature that investigates the effectiveness of incentives in general. Some laboratory and field experiments document that (small) incentives backfire and thus challenge the belief in the effectiveness of incentives.[22] Frey and Jegen (2001) stress that introducing incentives has two countervailing effects: besides the standard relative price effect, incentives may crowd out intrinsic motivation. With small incentives the relative price effect is small and the latter, counterproductive effect may dominate.

The paper proceeds as follows. Section 2 presents the experimental design and procedure, section 3 the behavioral predictions. Across and within subjects results are

[21] Torgler (2002) reviews the experimental literature on tax evasion and concludes that evidence on the effectiveness of deterrent incentives is rather mixed (p.662).

[22] Bowles (2007), Fehr and Falk (2002) and Frey and Jegen (2001) survey the economic literature on crowding out of intrinsic motivation. The origins of this literature are in psychology, see for example Deci (1971) and Lepper, Greene and Nisbett (1973). Deci, Koestner and Ryan (1999) provide a meta-analysis of more than 100 psychological studies on the effect of extrinsic rewards on intrinsic motivation.

Figure 2.1: Structure of the game



summarized and discussed in section 4. In section 5 we check the robustness of our results by presenting results from sessions with a moral frame. Section 6 concludes.

2.2 Experimental design and procedure

Consider one of the simplest possible games of stealing with two agents, A and B. Agent A is initially endowed with w_A, and agent B is initially endowed with w_B, where $w_A > w_B$.[23] While agent A is passive, agent B can take any amount $x \in [0, w_A]$ from agent A's endowment. If B does not take anything, i.e. $x = 0$, agents A and B both receive their initial endowments w_A and w_B. If B takes a strictly positive amount ($x > 0$), with probability $(1 - p) \in [0, 1]$ the taken amount x is indeed transferred from A to B; with probability p, however, x is not transferred and, on top of that, agent B has to pay a fixed fine f. We use a fixed fine that is independent of the taken amount to keep the design as simple as possible. The structure of the game is summarized in Figure 2.1.

Since we focus on pure incentive effects on B's behavior we vary the detection

[23] $w_A > w_B$ allows to distinguish between subjects who have a preference for fair (equal) outcomes and subjects who simply do not want to take anything in treatment T1.

probability p and the fine f across different treatments and fix w_A and w_B at levels 90 and 50, respectively. Table 2.1 presents the treatments.

Table 2.1: Treatments

Treatment	p	f	B's expected payoff given x = 0	B's expected payoff given x = 90	Level of incentives
T1	0.0	0	50	140	zero
T2	0.6	6	50	82.4	small
T3	0.5	25	50	82.5	small
T4	0.6	20	50	74	intermediate
T5	0.7	40	50	49	high
T6	0.8	40	50	36	very high

Treatment T1, our benchmark treatment, implements no deterrent incentives. It is simply the mirror image of a dictator game.[24] In all other treatments strictly positive detection probabilities and fines are implemented. We categorize the intensity of incentives according to B's expected payoff from taking agent A's whole initial endowment. As Table 2.1 shows the level of incentives is weakly increasing in treatment order. In treatments T2, T3 and T4 taking everything pays off in expectation. Treatment T5 is characterized by a combination of p and f such that taking the maximum possible amount generates about the same expected payoff as taking nothing. In treatment T6 the expected payoff from taking everything is strictly smaller than the one of taking nothing. Since the same intensity of incentives is implemented by different values of p and f in treatments T2 and T3, we can analyze whether p and f are interchangeable instruments in achieving deterrence, at least for this particular level of incentives.

Each experimental session consisted of three parts: two of the treatments described above and a dictator game.[25] After these three parts participants filled out a questionnaire eliciting data on their age, sex and subject of studies. We used a paid Holt and Laury (2002) procedure to elicit subjects' risk preferences.[26] The conducted sessions

[24] Here, subjects can decide how much to take away from (instead of to give to) another agent in a purely distributional context without any strategic considerations.

[25] In the dictator game, the dictator could give any amount of his initial endowment of 90 to a passive agent with an initial endowment of 50. The chosen amount may indicate the dictator's aversion to advantageous inequity. However, the donated amount might be affected by the treatments played in part 1 and part 2.

[26] The translated table and a brief report on the observed levels of risk aversion can be found in appendix 2.7.2.

are presented in Table 2.2.

Table 2.2: Session plan

Session	Part 1	Part 2	Part 3	Questionnaire	Number of participants
T1T3	T1	T3	DG	Yes	38
T3T1	T3	T1	DG	Yes	38
T2T3	T2	T3	DG	Yes	18
T3T2	T3	T2	DG	Yes	20
T2T4	T2	T4	DG	Yes	38
T4T2	T4	T2	DG	Yes	36
T5T6	T5	T6	DG	Yes	32
T6T5	T6	T5	DG	Yes	38

* includes a Holt and Laury (2002) table

At the beginning of each session, participants were told that one randomly picked part out of the three would be paid for all of them. After each part only the instructions for the following part were handed out. Subjects did not receive any feedback before the end of the experiment. They were matched according to a perfect stranger design, i.e. a couple matched once is never matched again in the following parts. Those subjects randomly chosen to be agents B in part 1 remained agents B in part 2 and were dictators in part 3. Consequently, passive subjects remained passive throughout all three parts of the session.[27]

This design offers the possibility to analyze the observed behavior in two different ways. First, we can compare behavior in part 1 across different treatments. This is the cleanest comparison because individual behavior in part 1 is not influenced by any preplay. Second, we can analyze how agents B adapt their behavior to the change in incentives from part 1 to part 2. Since the structure of the model is very simple we assume that a change in behavior from part 1 to part 2 is stimulated by the change of incentives rather than learning.

Our experimental sessions were run in November 2006 and March 2007 at the experimental laboratory of the SFB 504 in Mannheim, Germany. 258 students of the Universities of Mannheim and Heidelberg participated in the experiment. Subjects were randomly assigned to sessions and could take part only once. Except for the

[27] To keep passive subjects busy we asked them how they would decide if they were agent B.

robustness check, the sessions were framed neutrally[28] and lasted about 40 minutes. Subjects did not receive a show-up fee[29] and earned 12.34 € on average.

2.3 Behavioral predictions and hypotheses

2.3.1 Behavioral predictions

How the intensity of incentives affects B's transfer decision depends on the specific form of his utility function.

Model 1: The self-interest model

The standard neoclassical approach assumes that all people are selfish, i.e. their utility function U depends on their own material payoff m only and is increasing in m. With these assumptions the deterrence hypothesis holds, namely the optimal taken amount $x^*(p, f)$ is monotonically (weakly) decreasing in p and in f. Due to the fixed fine f agent B who maximizes his expected utility either takes as much as possible (w_A) or nothing. This depends on the relative sizes of p, f, w_A, w_B and on the level of risk aversion. B's optimal taken amount is

$$x^*(p,f) \in \begin{cases} \{0\} & if \quad p > \frac{U(w_A+w_B)-U(w_B)}{U(w_A+w_B)-U(w_B-f)} \\ \{0, w_A\} & if \quad p = \frac{U(w_A+w_B)-U(w_B)}{U(w_A+w_B)-U(w_B-f)} \\ \{w_A\} & if \quad p < \frac{U(w_A+w_B)-U(w_B)}{U(w_A+w_B)-U(w_B-f)} \end{cases} .$$

The higher p and/or the higher f, the less attractive it is to take everything. For sufficiently high values of p and f agent B does not take anything. A higher level of risk aversion[30] reduces $\frac{U(w_A+w_B)-U(w_B)}{U(w_A+w_B)-U(w_B-f)}$ ceteris paribus, and thus the set of p, f, w_A, w_B combinations for which taking everything is optimal.

In our two-agent setup with unequal initial endowments it is likely that agents B compare their own payoff to that of agent A. Thus, our setup is a context in which it

[28] Translated instructions for players B can be found in appendix 2.7.1.

[29] Six subjects did not earn anything in the randomly selected part and in the Holt and Laury (2002) table.

[30] Consider the Arrow-Pratt measure of relative risk aversion $-\frac{m*U''(m)}{U'(m)}$, for example.

is very plausible to consider preferences that reflect concerns for fairness.

Model 2: A model of fairness concerns regarding final outcomes

Models of fairness concerns regarding final outcomes (e.g. Fehr and Schmidt, 1999, Bolton and Ockenfels, 2000) assume that an agent's utility function $\widetilde{U}$ is increasing in the agent's own material payoff m, but decreasing in payoff inequality $|m - y|$ where y is the material payoff of the other agent.

In these models, the deterrence hypothesis still holds if there exists a unique optimal taken amount $x^*(p = 0, f = 0)$ that maximizes agent B's expected utility for $p = 0$ and $f = 0$. Due to the fixed fine f agent B who maximizes his expected utility either takes an amount which is optimal for no incentives ($x^*(p = 0, f = 0)$) or nothing. For relatively low values of p and f agent B takes $x^*(p = 0, f = 0)$ that may be smaller than w_A. For relatively strong incentives agent B is deterred and takes nothing.

The reason is that agents cannot trade off payoffs from different states, in our context payoffs if their transfer is detected and payoffs if their transfer is not detected. Assume there exists a unique optimal transfer amount for $p = 0$ and $f = 0$, $x^*(p = 0, f = 0)$. Then, $x^*(p \geq 0, f \geq 0)$ cannot be strictly larger than $x^*(p = 0, f = 0)$: If B's transfer is not detected, utility $\widetilde{U}$ is maximized at $x^*(p = 0, f = 0)$. If B's transfer is detected, utility $\widetilde{U}$ is the same for any strictly positive transfer amount and largest for $x = 0$. Analogously, taking any strictly positive, but strictly smaller amount than $x^*(p = 0, f = 0)$ yields less expected utility than taking $x^*(p = 0, f = 0)$ and, therefore, cannot be optimal.

Model 3: A model of fairness concerns regarding expected outcomes

Models of fairness concerns regarding expected outcomes (e.g. Trautmann, 2007) assume that an agent's utility function $\widehat{U}$ is increasing in the agent's own material payoff m and decreasing in the absolute difference between own expected payoff m^e and the other agent's expected payoff y^e, $|m^e - y^e|$.

If $|m^e - y^e|$ directly enters the utility function, the deterrence hypothesis may not hold anymore, i.e. $x^*(p, f)$ may be strictly *increasing* in p and/or f for some values of $p \geq 0$ and $f \geq 0$.

The reason is that agents can trade off payoffs from different states, e.g. advantageous inequity in material payoffs if B's taking is not detected can compensate disadvantageous inequity in material payoffs if B's taking is detected. As an illustration consider the following utility function $\widehat{U} = m - \beta * max\{m^e - y^e, 0\}$ with $m^e = (1-p) * (w_B + x) + p * (w_B - f)$ and $y^e = (1-p) * (w_A - x) + p * w_A$. If $\beta > \frac{1}{2}$, agent B who maximizes expected utility tries to perfectly equate m^e and y^e by choosing x. Hence, agent B takes more the higher p and/or the higher f. However, since the transfer amount x is bounded above by w_A, agent B is deterred from taking if p and f are sufficiently high.[31]

Model 4: A model of fairness concerns regarding final outcomes that are crowded out by incentives

The literature on crowding out of intrinsic motivation through extrinsic incentives uses the term "intrinsic motivation" very broadly. It may apply to social preferences or fairness concerns as well. In our context, crowding out implies that agents' fairness concerns are decreasing in the intensity of deterrent incentives. Formally, this assumption can be captured by the following utility function:

$$V = \lambda(p, f) * U(m) + [1 - \lambda(p, f)] * \widetilde{U}(m, |m - y|),$$

where as above $U(m)$ represents utility of a selfish agent and $\widetilde{U}(m, |m-y|)$ utility of an agent with fairness concerns regarding final outcomes. The core of the crowding out assumption is that $\lambda(p, f) \in [0, 1]$, the weight of $U(m)$, is (weakly) increasing in p and in f.

With these assumptions, there may be ranges of p, f combinations such that the optimal transfer amount $x^*(p, f)$ *increases* strictly in p and/or f. This is not consistent with the deterrence hypothesis.

The intuition is that for small values of p and f agent B is relatively fair-minded and takes only a small amount. If, however, p and f are high, agent B is rather selfish and takes a larger amount. Only if the levels of p and f are very high such that both selfish and fair-minded agents are deterred, agent B does not take anything.

[31] Consider $\widehat{U}$ and assume that p and f are so high that taking everything would yield $m^e < y^e$ with $m^e < w_B$. In this case taking nothing is optimal.

If we should observe crowding out caused by deterrent incentives, our data may be very useful to contribute evidence to two aspects of crowding on which the verdict is still out.

(i) Continuity of crowding out

$\lambda(p, f) \in [0, 1]$ may increase continuously or discontinuously in p and in f. Even if $\lambda(p, f)$ increases continuously, the taken amount $x^*(p, f)$ may increase discontinuously in p and in f for some $\widetilde{U}(m, |m - y|)$.

The empirical results of Gneezy and Rustichini (2000b) and Gneezy (2003) suggest discontinuous crowding out. Frey and Oberholzer-Gee (1997), however, explain their data by assuming continuous crowding out.

We have evidence for continuous crowding out if we observe subjects who take less than the maximal transfer amount w_A in a second treatment in which incentives have been introduced or increased compared to the first treatment.

(ii) Hysteresis

Extrinsic incentives may crowd out fairness concerns lastingly. As a consequence the crowding out effect of an increase in incentives is larger than the crowding in effect of the *subsequent* decrease in incentives that reverses the increase in incentives by size.

Some studies (e.g. Irlenbusch and Sliwka, 2005, Gneezy and Rustichini, 2000, Gächter, Königstein and Kessler, 2005) find evidence for hysteresis, i.e. evidence that incentives crowd out fairness concerns lastingly.

To give an example, if subjects take more in treatment T2 with small incentives if T2 is played second after treatment T4 with intermediate incentives than if it is played first we have evidence for hysteresis.

If we find backfiring of deterrent incentives plus hysteresis in our data, our results can only be explained by crowding out of fairness concerns and not by a model of fairness concerns regarding expected outcomes. Hence, hysteresis might be a mean to distinguish between the two possible models (model 3 and 4) that can explain backfiring incentives.

2.3.2 Hypotheses

The predictions of the various models differ. However, all four models presented above support hypothesis 1.

Hypothesis 1: Deterrence by strong incentives

Sufficiently high values of the detection probability p and the fixed fine f deter agents from taking. This range is larger, the more risk averse an agent is.

The threshold of incentives that effectively deter may vary by subject. A risk neutral or risk averse selfish agent will abstain from taking in treatments T5 and T6. A risk neutral or risk averse agent with standard Fehr-Schmidt (1999) fairness preferences may even abstain from taking in treatment T4. A subject with fairness concerns regarding expected outcomes may only abstain from taking if $p = 1$ and $f > 0$.

In contrast to hypothesis 1, only the self-interest model and the model of fairness concerns regarding final outcomes necessarily imply hypothesis 2, but not the model of fairness concerns regarding expected outcomes or the model of fairness concerns that are crowded out by deterrent incentives.

Hypothesis 2: Deterrence hypothesis

The taken amount x is monotonically (weakly) decreasing in the detection probability p and the fixed fine f.

Hypothesis 2 implies that the average taken amount x should be (weakly) decreasing from treatments T1 to T6.

2.4 Results

First, we will compare the different treatments in part 1, i.e. do an across subjects analysis. This has the advantage that behavior is not influenced by any preplay. However, we cannot draw any conclusion whether the observed across subjects phenomena show up in individual level data and whether hysteresis occurs. Second, we will address these issues by comparing behavior in part 1 and 2 with a focus on within subject analysis.

2.4.1 Comparison of treatments in part 1

Summary statistics

Benchmark treatment

The experimental data in treatment T1 show how much people take in the absence of deterrent incentives. The upper left panel of Figure 2.2[32] summarizes the distribution of the taken amount x in the benchmark treatment.

Since treatment T1 is the mirror image of a dictator game, we can compare behavior in T1 with standard results of dictator games as for example Forsythe et al. (1994). In line with their paper, we can identify two types of agents: selfish agents and fair-minded agents. In their benchmark treatment (the paid dictator game conducted in April with a pie of 5 \$) about 45 % of subjects are "pure gamesmen" who do not give anything, and the rest gives a strictly positive amount. These types of agents correspond remarkably well to the 47 % (52.5 %) of selfish subjects in treatment T1 who take everything (between 80 and 90), and the rest who takes a strictly positive amount below 90 (80).

To summarize, we have two types of agents: slightly less than 50 % of our subjects have standard preferences while a bit more than 50 % have fairness concerns. As model 2 shows fairness concerns are not necessarily a reason why Becker's deterrence hypothesis might fail. But it may fail if fairness concerns refer to expected outcomes or if fairness concerns are crowded out by deterrent incentives. To figure out whether this is the case we have a closer look at the treatments with deterrent incentives.

Treatments with deterrent incentives

Figure 2.3 summarizes the average taken amount x per treatment.

Treatments are ordered by the intensity of deterrent incentives, i.e. the combined effect of detection probability p and fine f (compare Table 2.1). The average taken

[32] In Figure 2.2 intervals with $y < 90$ are the union of all points $x \in [y, y + 5)$ with y denoted at the horizontal axis. The last interval contains all $x = 90$.

Figure 2.2: Distributions of taken amounts (in intervals of size 5)

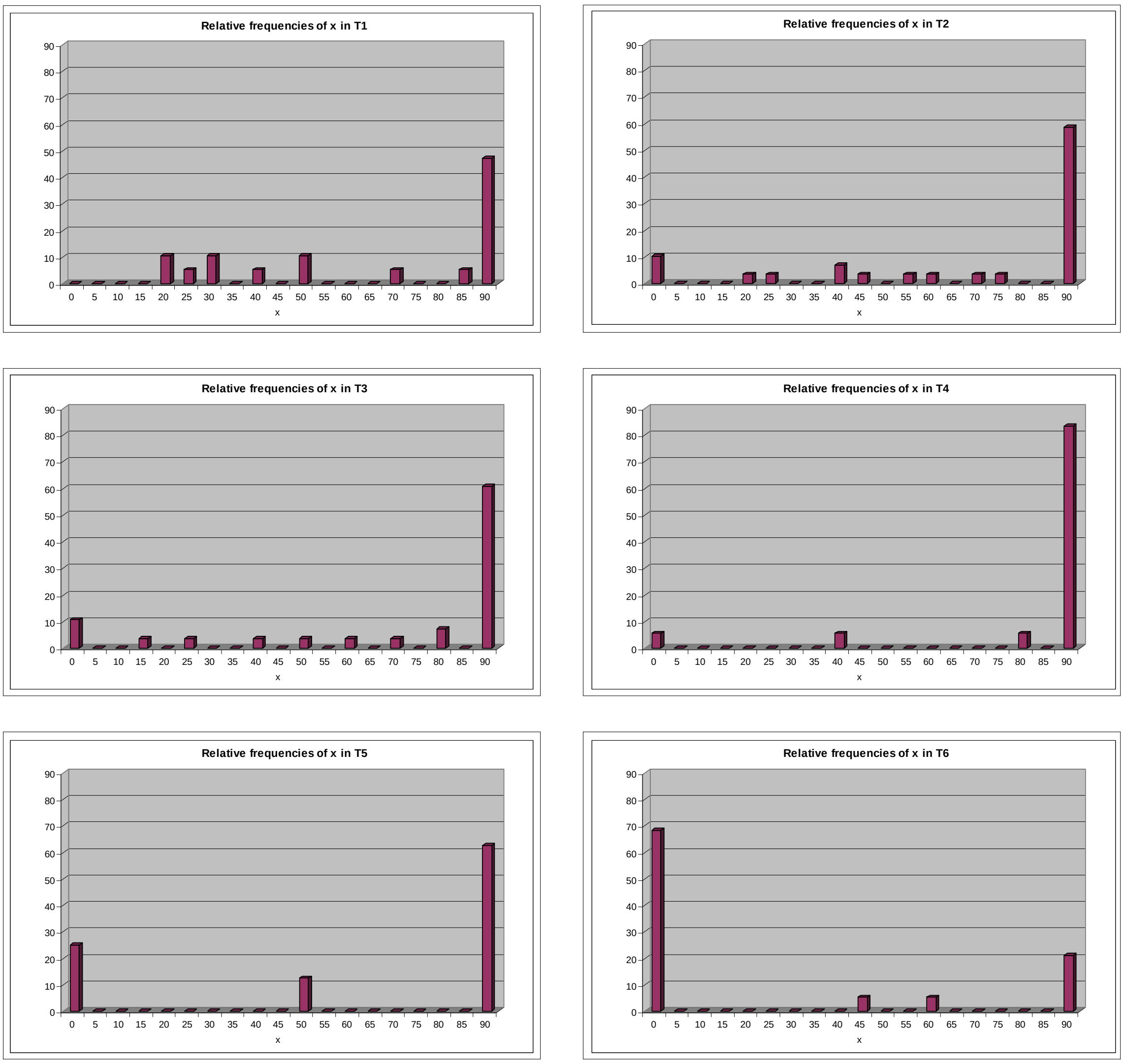
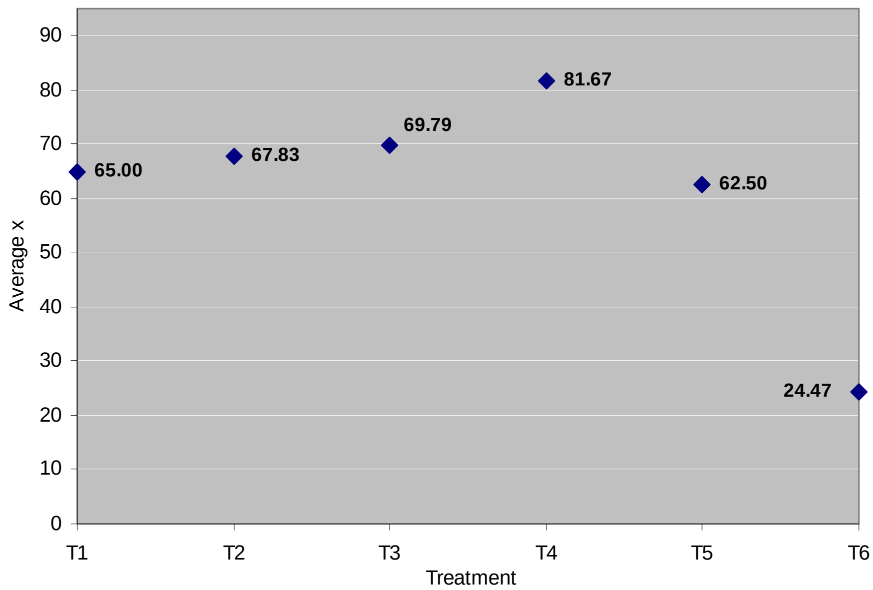


Figure 2.3: Average taken amount by treatment



amount x increases in the range of no, small and intermediate incentives (from T1 to T4) while it decreases in the range of strong and very strong incentives (T5 and T6). Hence, the relationship between the average taken amount and the intensity of deterrent incentives is rather inverted-U shaped than monotonically decreasing.

Figure 2.2 shows that the fraction of subjects taking everything increases by treatment from T1 to T4. In treatment T4, it peaks at a value of more than 80 % which is considerably higher than the corresponding 47 % in the absence of any incentives in treatment T1. From treatment T5 onwards this fraction decreases.

The share of subjects not taking anything monotonically increases in the level of incentives. It is moderate with no, small and intermediate incentives ($\leq$ 10 %), quite substantial with high incentives (about 25 %) and largest with very strong incentives (nearly 70 %).

There are always subjects taking interior values of their strategy set, most so in the benchmark treatment. The share of these subjects decreases in the intensity of incentives. Moreover, the average transfer amount conditional on interior values increases by treatment from T1 to T4.

Compared to the benchmark treatment deterrent incentives shift mass to the borders of the support. We observe both backfiring of small incentives and deterrence at the same time.[33] Small and intermediate incentives move mass predominately towards the upper border which stands in sharp contrast to the deterrence hypothesis, but is consistent with a model of fairness concerns regarding expected outcomes or a model of fairness concerns that are crowded out by deterrent incentives. Strong and very strong incentives move mass exactly to the lower border which is consistent with hypothesis 1. Since the results of treatments T2 and T3 are very similar, detection probability and fine seem to be interchangeable instruments.

Analysis of hypotheses

A Kruskal-Wallis test on behavior in part 1 documents significant ($p < 0.01$) treatment effects. In order to identify and characterize the significant differences we run pair wise Mann-Whitney tests. The one-sided p-values are recorded in Table 2.3.

Table 2.3: Pair wise treatment comparisons (Mann-Whitney tests)

	T2	T3	T4	T5	T6
T1	0.287	0.234	0.015	0.400	< 0.001
T2		0.408	0.040	0.447	< 0.001
T3			0.058	0.390	< 0.001
T4				0.071	< 0.001
T5					0.005

*: We report one-sided p-values.

In treatment T6, agents take significantly ($p < 0.01$) less than in any other treatment. This is consistent with hypotheses 1 and 2. However, contradictory to hypothesis 2, the deterrence hypothesis, agents take significantly more in treatment T4 than in treatments T1 ($p < 0.05$), T2 ($p < 0.05$) and T3 ($p = 0.058$).[34] There is no significant difference in behavior in treatments T2 and T3.

[33] In an experiment on corruption that uses probabilistic incentives as we do, Schulze and Frank (2003) observe a similar pattern in their data.

[34] One-sided Kolmogorov-Smirnov tests and χ^2-tests based on a grouping of subjects according to whether they are deterred, try to roughly equate payoffs (take between 15 and 29 units), show some fairness concerns (take between 30 and 79 units) or are selfish (take between 80 and 90 units) largely confirm the results of the Mann-Whitney tests presented here. In particular, subjects always take significantly more in treatment T4 than in T1.

In order to account for individual characteristics when comparing treatments we estimate two specifications whose marginal effects are presented in Table 2.4.

Table 2.4: Regression results (OLS and Tobit)

dependent variable: taken amount x explanatory variables	OLS		Tobit	
Intercept	57.03	***	94.54	***
Sex (1 if male, 0 else)	12.14	*	28.08	
Risk aversion (1 if risk averse, 0 else)	-14.55	**	-57.16	**
Economist (1 if economist, 0 else)	10.05	*	30.48	
DG (donated amount in part 3)	-0.12		-0.48	
T2	10.23		33.81	
T3	9.04		27.32	
T4	18.38	**	87.65	**
T5	-7.38		-20.79	
T6	-42.74	***	-132.64	***
N	129		129	
(Pseudo) R^2	0.305		0.075	

comments: Ti = 1 if treatment = Ti, 0 else
*, **, *** significant at 10, 5, 1 % level based on robust standard errors

First, we regress the taken amount on individual characteristics and treatment dummies using OLS estimation. Second, we address the fact that the taken amount is truncated and estimate a Tobit specification with the same explanatory variables.

In both estimations the treatment dummy for T4 is significantly positive ($p < 0.05$), the treatment dummy for T6 is significantly negative ($p < 0.05$) and the treatment dummies for T2 and T3 are not significantly different from each other.[35] Hence, our test results are very robust to including control variables. Risk aversion has a significantly negative effect ($p < 0.05$) on the taken amount in the both specifications (as subjects with a high level of risk aversion are more likely to be deterred).

Given our test and estimation results we do not reject hypothesis 1, but we do reject hypothesis 2, the deterrence hypothesis.

Result 1: Deterrence by strong incentives

Very strong incentives as in treatment T6 significantly reduce the taken amount. On average, risk averse agents take significantly less.

[35] Adding interaction effects of the dummy for risk aversion with the treatment dummies in the OLS estimation does not change any of these results.

Result 2: Backfiring of small incentives

Deterrent incentives do not monotonically (weakly) decrease the average taken amount. Intermediate incentives in treatment T4 significantly increase the average taken amount.

Result 3: Interchangeability of detection probability and fine

We do not find any significant differences between treatments T2 and T3. In that sense, detection probability p and fine f seem to be interchangeable policy instruments.

In sum, these results are consistent with the predictions of the model of fairness concerns regarding expected outcomes or the model of fairness concerns that are crowded out by deterrent incentives.

2.4.2 Comparison of behavior in part 1 and 2

Up to now we have compared different treatments across *different subjects*. In contrast to Becker's deterrence hypothesis our results so far show that intermediate incentives backfire. Crowding out of fairness concerns or a model of fairness concerns regarding expected outcomes are explanations for this phenomenon. Since each subject sequentially participated in two different treatments, we can further analyze how *the same individuals* react to a change of deterrent incentives.[36] Sessions in which we increase incentives allow to analyze (i) whether backfiring of small and intermediate incentives is observed on an individual level, and (ii) whether backfiring is a continuous or discontinuous process. Sessions in which we decrease incentives enable us to check whether we observe hysteresis. Sessions with incentives of constant intensity indicate whether p and f are interchangeable instruments at an individual level.

Backfiring of incentives at an individual level

In three different sessions, we increase the intensity of incentives from part 1 to part 2: in session T1T3 from no to small incentives, in session T2T4 from small to inter-

[36] Since subjects do not get any feedback after part 1, behavioral effects cannot be triggered by the realization of punishment.

mediate incentives, in session T5T6 from strong to very strong incentives. Figure 2.4 summarizes how subjects behave in part 2 conditional on whether they acted selfishly ($x = 90$), acted fair-mindedly ($0 < x < 90$) or were deterred ($x = 0$) in part 1.

Since the benchmark treatment was played in the first part of session T1T3, we can identify about 47 % of subjects with standard preferences. All except one take everything in part 2 again. 53 % of all subjects take an amount strictly less than everything in part 1. About a third of them increase the taken amount x to a level smaller than 90, a fifth is switches to taking everything in part 2 and another fifth keeps x constant. Hence, for 50 % of fair subjects small incentives seem to strictly backfire. Only one selfish and one fair-minded subject are deterred by small incentives.

In session T2T4, about 63 % of subjects take everything already in part 1. We cannot distinguish whether they have standard preferences or whether they have fairness concerns that are crowded out by the small incentives in part 1. Again, the majority of these subjects is not deterred and keeps taking everything in part 2. The share of subjects taking intermediate amounts in part 1 is considerably smaller than in session T1T3. For 20 % of these subjects the increase of incentives completely backfires. The majority, however, is deterred. Note that a moderate fraction of deterrence can already be found in part 1.[37]

In session T5T6, 62.5 % of subjects still take everything in part 1. More than two thirds of them are deterred by the increase of incentives though. 25 % of all subjects are deterred in part 1 and stay deterred in part 2. Only 12.5 % of subjects take a strictly positive amount below 90 in part 1. Half of them are deterred in the second part.

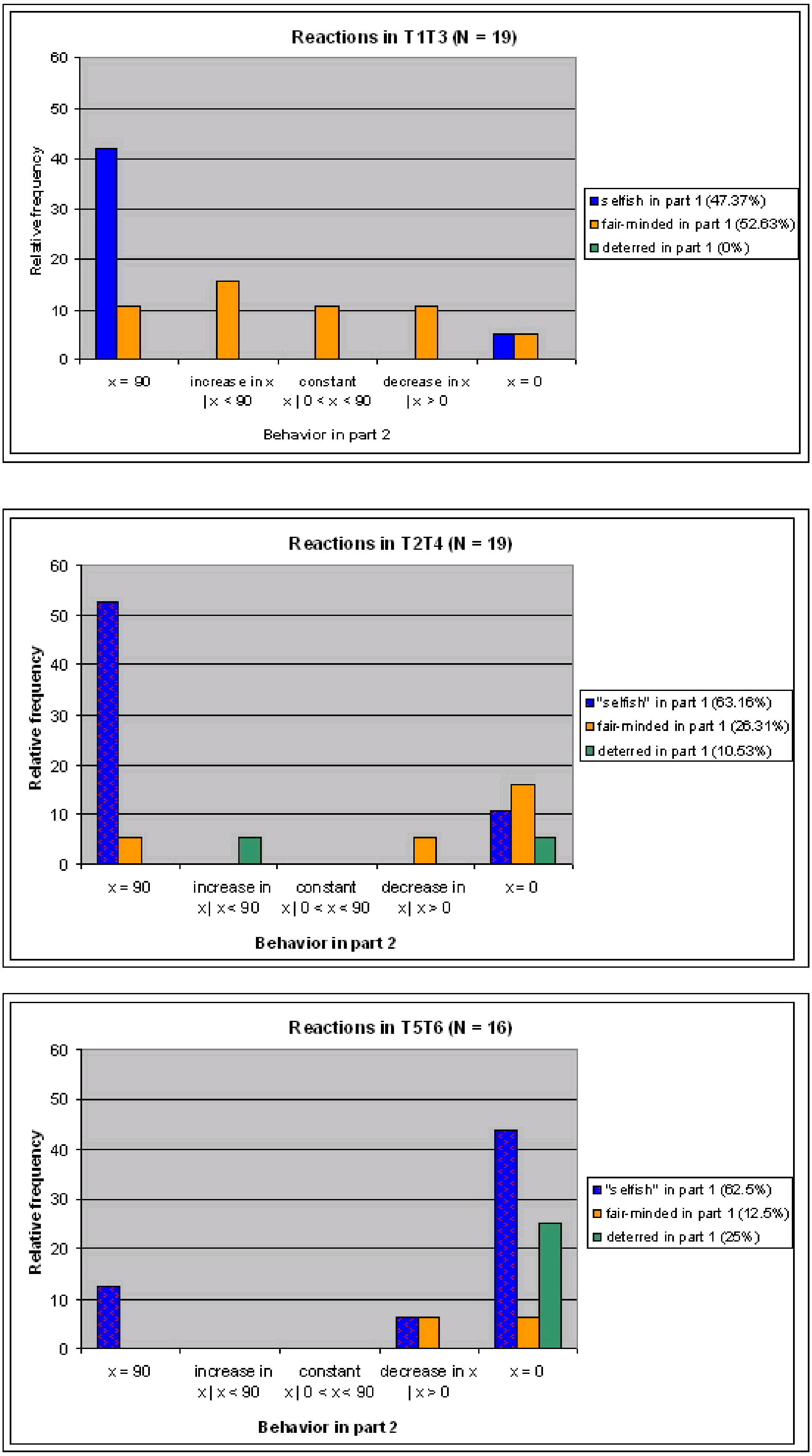
These observations can be summarized in the following two results:

Result 4: Backfiring of incentives on an individual level

Subjects seem to be heterogeneous. There are selfish agents for whom the deterrence hypothesis holds. However, there are also fair-minded agents for whom small and

[37] In sessions T1T3 and T2T4 none of the proposed models can explain the behavior of subjects who react to increased incentives by decreasing the taken amount to a level larger than 0 or increasing it from 0 to a strictly positive amount.

Figure 2.4: Reactions to an increase in the intensity of incentives



intermediate incentives backfire. Independent of the type of agent strong incentives deter.

Result 5: Continuous and discontinuous backfiring of incentives

We find evidence for both continuous and discontinuous backfiring of incentives.

Hysteresis

Whether hysteresis (lasting crowding out) is present in our data can be seen by comparing behavior of a given treatment played in part 1 with behavior of the same treatment played in part 2 after a part 1 treatment with stronger incentives. Hysteresis implies that we should observe order effects for treatments T1, T2 and T5 because they are played second in sessions in which we decrease deterrent incentives (T3T1, T4T2, T6T5). Table 2.5 records two-sided p-values of pair wise Mann-Whitney tests that compare the *same* treatment played in *different* parts of a session.[38]

Table 2.5: Order effects (Mann-Whitney tests)

Treatment	played first in	played second in	p-value (two sided)
T1	T1T3	T3T1	0.082
T2	T2T3 T2T4	T3T2 T4T2	0.099
T3	T3T1 T3T2	T1T3 T2T3	0.676
T4	T4T2	T2T4	0.061
T5	T5T6	T6T5	0.014
T6	T6T5	T5T6	0.617

As Table 2.5 indicates we indeed observe order effects in treatments T1, T2 and T5. Subjects in T1 take significantly ($p < 0.05$) more when it is played after T3 (81.3 instead of 65.0 units on average). Preplay in T3 with small incentives increases the average taken amount in treatment T1 that does not implement any incentives. Similarly, the average taken amount in T2 is significantly ($p < 0.05$) higher when it is played

[38]Treatments T2 and T3 are played second in two different sessions. Since the observations from the different second parts are not significantly different (p=0.71 and p=0.34, respectively according to two-sided Mann-Whitney tests) for different sessions, we do not report each session comparison separately.

second (after a harsher or a constant intensity of incentives) than first. Both results are consistent with a model of lasting crowding out but cannot be reconciled with the predictions of a model of fairness concerns regarding expected outcomes. In contrast, in T5 with strong deterrent incentives subjects take significantly ($p < 0.05$) less when it is played after T6. Preplay in T6 with very strong incentives seems to increase deterrence in treatment T5. This is inconsistent with both a model of lasting crowding out and a model of fairness concerns regarding expected outcomes (if fairness concerns imply that fair subjects take weakly less in T5 than selfish subjects for a given level of risk aversion). However, we do not expect that preplay in T6 significantly increases the taken amount x in treatment T5 as risk neutral or risk averse selfish agents are deterred anyway.

Result 6: Hysteresis

Small and intermediate incentives have a lasting effect. They still backfire when incentives are decreased or even removed in the following period.

Since we observe hysteresis, a model of crowding out of fairness concerns explains our data better than a model of fairness concerns regarding expected outcomes. Hysteresis also underlines how costly extrinsic incentives are. In addition to the effect incentives have in the current period they may also influence behavior in future periods. Also strong incentives could potentially backfire by crowding out fairness concerns in future periods in which incentives are smaller.

In treatments with an increase in incentives there are no significant order effects for treatments T3 and T6, but subjects in treatment T4 take significantly ($p < 0.05$) less when it is played in part 2 after treatment T2 than when treatment T4 is played in part 1.

Substitutability of detection probability and fine

Since treatments T2 and T3 have the same intensity of deterrent incentives implemented by different levels of detection probability p and fine f we can test - at least for this specific level of incentives - whether these two instruments are interchangeable. We have already seen that the treatments T2 and T3 do not differ significantly in across subjects data when they are played in part 1 (result 3). Our within subject data in Figures 2.5 confirm this result.

In session T3T2 only a single subject is apart from the 45° line. 6 subjects keep taking everything, 2 keep taking the same intermediate amount. In session T2T3 7 out of 10 subjects do not change their behavior.

Result 7: Interchangeability of detection probability and fine at an individual level

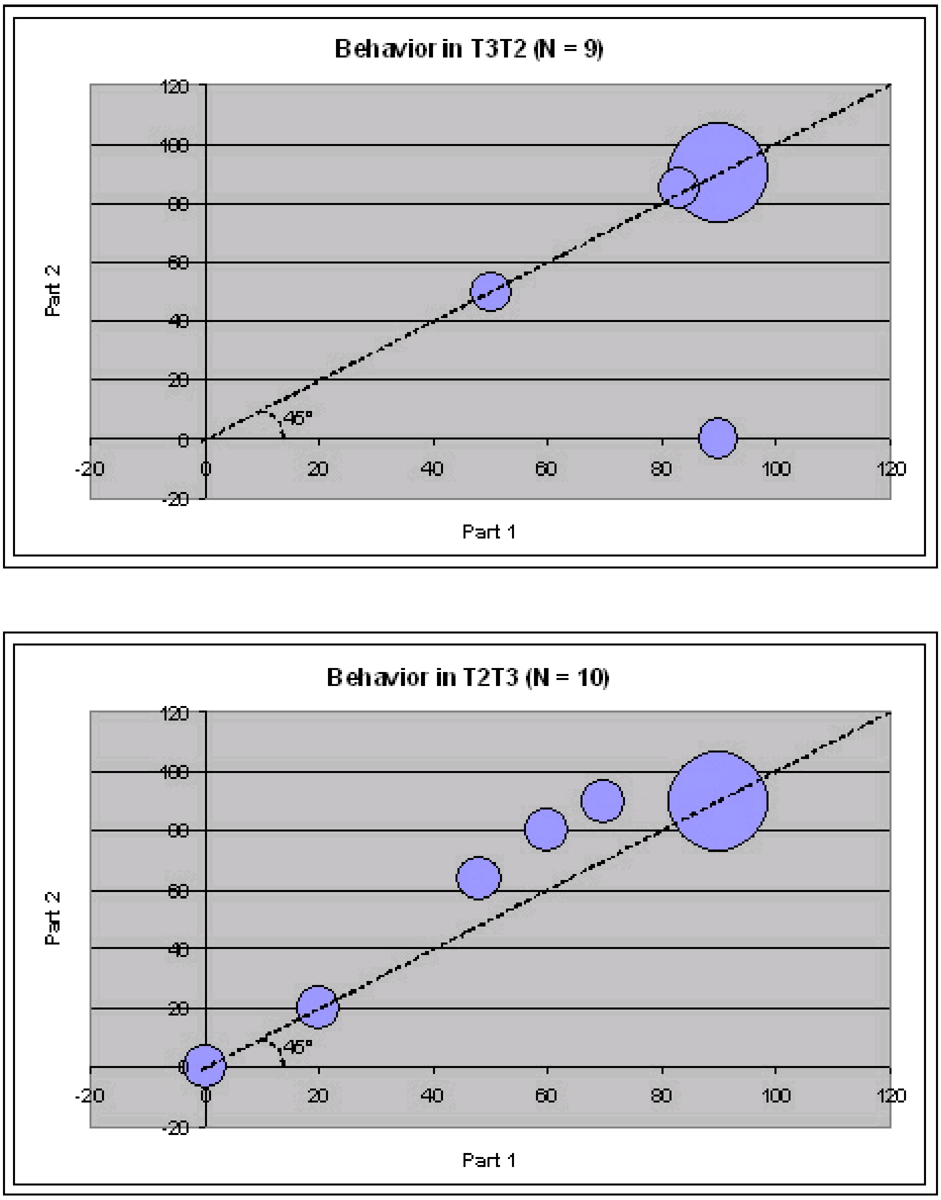
Within subject data confirm result 3 that p and f seem to be interchangeable instruments.

2.5 Robustness check - Framing

So far we have presented results from neutrally framed experiments. This is a valid approach to testing Becker's (1986) deterrence hypothesis that relies on pure incentive effects. While a non-neutral frame may ceteris paribus affect the taken amount (e.g. due to additional moral costs) comparative statics should remain unchanged. For any given (neutral or non-neutral) frame the deterrence hypothesis predicts the taken amount to be monotonically decreasing in detection probability and fine. In contrast, it is not clear whether a non-neutral frame interacts with incentives in the model of fairness concerns that are crowded out by incentives which fits our data best. While neutrally framed incentives crowd out fairness concerns, this may not necessarily be the case for incentives that are combined with a strong moral connotation.

In real life deterrent incentives often have a moral connotation and policy makers may try to make use of that. This is why we run two additional morally framed sessions and have a look at whether a non-neutral, moral frame will change our results. In the morally framed sessions the "transfer decision" was labeled as "stealing" if $x > 0$ and the fixed fine f was called "penalty" instead of "minus points". Apart from these two different labels the neutrally and morally framed sessions were conducted in a completely identical manner. In order to check whether framing affects behavior in

Figure 2.5: Reactions to a change in incentives keeping their intensity constant



the absence of incentives we run a framed version of treatment T1 (T1f). To analyze whether framing and incentives interact we run a framed version of treatment T4 (T4f).[39] 38 subjects participated in session T1fT4f, 32 subjects in session T4fT1f.

The results in the framed and neutral treatments are similar. There is no significant framing effect in part 1 in the absence of incentives, i.e. in T1 and T1f (two-sided Mann-Whitney test: $p > 0.5$). In contrast, subjects take more in treatment T4 than in treatment T4f in part 1 (two-sided Mann-Whitney test: $p = 0.075$). However, the within subject results document a substantial degree of crowding out: when incentives are introduced in part 2 of session T1fT4f more than 30 % of individuals flip from intermediate amounts to taking everything. This parallels the results obtained in the neutrally framed sessions T1T3 and T2T4. In sum, we conclude that also with moral framing backfiring of intermediate incentives is a non-negligible phenomenon.

2.6 Conclusion

We have presented an experimental test of the deterrence hypothesis applied to the context of stealing. Our across subjects data reject the hypothesis that the average taken amount is monotonically decreasing in deterrent incentives. On average, subjects take most when intermediate incentives are present. Only very strong incentives deter.

Both our across subjects comparison of behavior in part 1 and our within subject comparison of behavior in part 1 and 2 reflect two different types of subjects. We identify 50 % selfish subjects whose behavior is consistent with the deterrence hypothesis and 50 % fair-minded subjects for whom intermediate incentives backfire. Since we observe hysteresis, a model of lasting crowding out of fairness concerns explains our data best.

We have contributed to the empirical literature on crowding out in various ways. First, we have established the existence of crowding out as a reaction to probabilistic

[39] We choose treatment T4 since the intensity of deterrent incentives in this treatment is (i) low enough not to deter the majority of individuals and (ii) high enough to potentially crowd out fairness concerns significantly.

incentives[40] and in a new domain, namely when incentives are set to deter criminal activities. Second, our within subject results provide further evidence for lasting crowding out as it is observed by Irlenbusch and Sliwka (2005), Gneezy and Rustichini (2000), and Gächter, Königstein and Kessler (2005). While lasting crowding out exists for many subjects, we have also observed some subjects whose fairness concerns are - at least partially - reestablished when incentives are reduced or removed completely. The circumstances under which crowding out is lasting remain a topic for future research. Third, our study has explicitly focused on the domain of small incentives that are especially important in real life[41]: we have run four out of six treatments with small incentives that according to standard neoclassical theory should not have any effect on a risk neutral subject's behavior. Thus, we have many observations to closely analyze whether crowding out is a continuous or discontinuous process. For example, the empirical results of Gneezy and Rustichini (2000b) and Gneezy (2003) suggest discontinuous crowding out. Frey and Oberholzer-Gee (1997), however, explain their data by assuming continuous crowding. Our within subject analysis finds evidence for both.

Furthermore, we observe crowding out of fairness concerns in a very simple setting which does not leave a lot of scope for the triggers of crowding out that are stressed in the theoretical literature. First, since incentives are set exogenously by the experimenter and are not endogenously determined by another subject, incentives cannot cause crowding out by signaling beliefs on other players' types (e.g. lack of trustworthiness) as in Ellingsen and Johannesson (2005) or by signaling norms as in Sliwka (2007). Second, an action's costs and benefits are common knowledge in our experiment. Thus, high incentives, for example, cannot cause crowding out by signaling high costs or low benefits of an action as in Bénabou and Tirole (2003). Last, in Bénabou and Tirole (2007) the presence of extrinsic incentives removes the possibility for agents to signal their own type (to other agents or to themselves). In our experiment, the initial en-

[40] To our knowledge the only other paper that documents crowding out of intrinsic motivation due to probabilistic incentives is Schulze and Frank (2003).

[41] In Germany, the clearance rate for thefts with (without) aggravating circumstances was 14 % (44 %) in 2005 (Polizeiliche Kriminalstatistik, 2005, Table 23). Andreoni, Erard and Feinstein (1998) present figures for tax evasion in the US: in 1995, the audit rate for individual tax return was only 1.7 %, the penalty for underpayment of taxes usually 20 % of the underpayment. Polinsky and Shavell (2000b) point out that in general the severity of punishment is quite low in relation to what potential offenders are capable to pay.

dowment of passive agents A is larger than the one of agents B such that agents B can signal their intrinsic motivation to behave fairly even if deterrent incentives are present. Kahneman and Tversky's (1986) argument that extrinsic incentives shift the context from an ethical and other-regarding to an instrumental and self-regarding one seems to be more adequate for our data. Similarly, our findings confirm those of Houser et al. (2007) who show that crowding out of intrinsic motivation is not only caused by the intentions that incentives signal but also by incentives per se.

What are the policy implications from our experimental study? Taking our data literally would imply to punish criminal behavior either hard or not at all to avoid backfiring of small incentives. Of course, the laboratory may abstract from social norms and stigmata that could be the driving forces behind punishment in reducing criminal behavior. Thus, we do not conclude that punishment does not work outside the laboratory. However, our data directly reject the deterrence hypothesis that relies on punishment whose effectiveness is caused by pure incentive effects. Our results show that if crime were a gamble - as economists generally argue and as we have modeled it in the laboratory - pure incentives may not work: Especially small and intermediate incentives backfire and may crowd out fairness concerns lastingly. Thus, to convincingly contribute to the discussion on how to efficiently deter crime economists should go beyond the standard deterrence hypothesis.

2.7 Appendix

2.7.1 Experimental sessions and instructions

The order of events during each experimental session was the following: Subjects were welcomed and randomly assigned a cubicle in the laboratory where they took their decisions in complete anonymity from the other participants. The random allocation to a cubicle also determined a subject's role in all three parts. Subjects were handed out the general instructions for the experiment as well as the instructions for part 1. After all subjects had read both instructions carefully and all remaining questions were answered, we proceeded to the decision stage of the first part. Part 2 and 3 were conducted in an analogous way. We finished each experimental session by letting subjects answer a questionnaire that asked for demographic characteristics and included a Holt and Laury (2002) table. This table was explained in detail in the questionnaire and it was highlighted that one randomly drawn decision from the table was paid out in addition to the earnings in the previous parts.

Instructions, the program and the questionnaire were originally written in German. The translated general instructions, the translated instructions of the neutrally framed treatment T4 in part 1 for agent B and the translated Holt and Laury (2002) table can be found in the following section. Instructions for part 2 and part 3 are as similar to part 1 as possible. For the framed treatments we used the expression "steal any integer amount between 0 and 90 from participant A" instead of "choose any integer amount between 0 and 90 that shall be transferred from participant A to you" and the term "minus a penalty of x points" instead of "minus an amount of x points".

Translated general instructions

General explanations concerning the experiment

Welcome to this experiment. You and the other participants are asked to make decisions. Your decisions as well as the decisions of the other participants will determine the result of the experiment. At the end of the experiment you will be paid **in cash** according to the **actual** result. So please read the instructions thoroughly and think about your decision carefully.

During the experiment you are not allowed to talk to the other participants, to use cell phones or to start any other programs on the computer. The neglect of these rules will lead to the immediate exclusion from the experiment and all payments. If you have any questions, please raise your hand. An experimenter will then come to your seat to answer your questions.

During the experiment we will talk about points instead of Euros. Your total income will therefore be calculated in points first. At the end of the experiment the total amount of points will be converted into Euros according to the following exchange rate:

1 point = 15 Cents.

The experiment consists of three **independent** parts in which you can accumulate points. Before each part only the instructions of this part will be handed out.

During the experiment neither you nor the other participants will receive any information on the course of the experiment (e.g. decisions of other participants or results of a particular part).

The results of each single part will be calculated only after all three parts will be finished. **Then, one of these three parts will be chosen randomly. At the end of the whole experiment only this part will be paid out in cash according to your decisions.**

Translated instructions of the neutrally framed treatment T4 in part 1

Part 1

In this part there are **participants in role A and participants in role B. You have been randomly assigned role B for this part. You will be randomly and anonymously matched to another participant in role A.** This random matching lasts only for this part. The matched participant will not be matched to you in the following two parts again. Neither before nor after the experiment will you receive any information about the identity of your matched participant. Likewise, your matched participant will not receive any information about your identity.

As participant B you have an initial endowment of 50 points. Participant A has an initial endowment of 90 points.

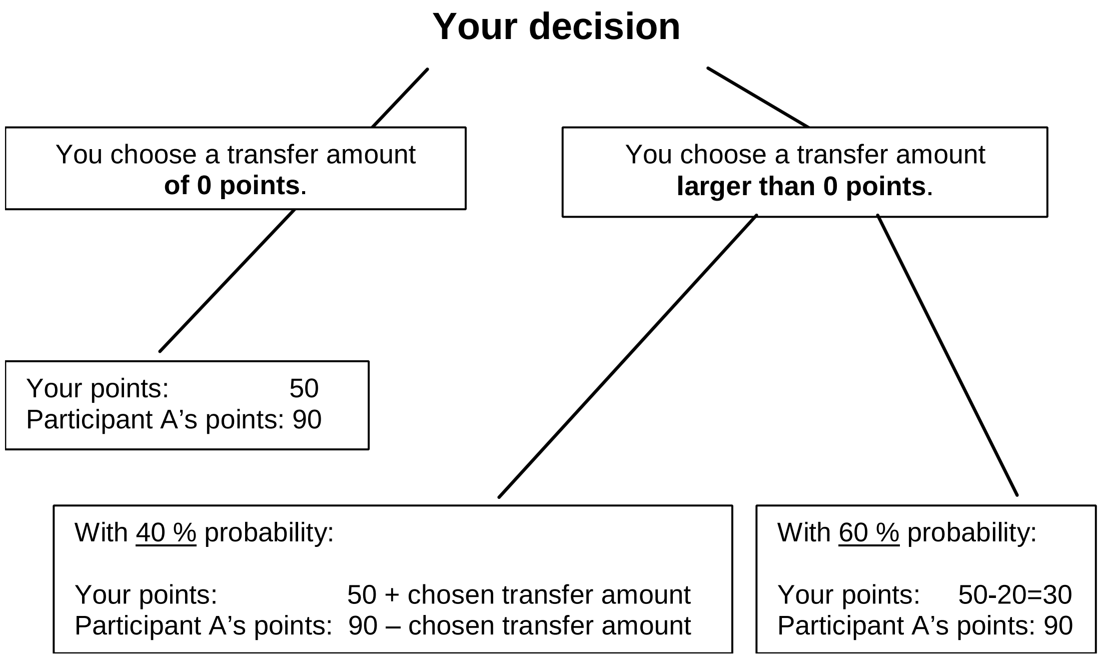
As a participant in role B you can choose any **integer amount** between 0 and 90 points (including 0 and 90) **which shall be transferred from participant A to you**. Participant A does not make any decision. In order to make your decision please enter your chosen amount on the corresponding computer screen and push the OK button.

- If you choose a transfer amount of 0 points, you will receive your initial endowment of 50 points, and participant A will receive his initial endowment of 90 points.
- If you choose a transfer amount larger than 0 points,
  - **with *40 %* probability you will receive your initial endowment of 50 points plus your chosen transfer amount and participant A will receive his initial endowment of 90 points minus your chosen transfer amount.**
  - **with *60 %* probability you will receive your initial endowment of 50 points minus an amount of 20 points, i.e. 30 points, and participant A will receive his initial endowment of 90 points.**

Example 1: You choose a transfer amount of 22 points. With *40 %* probability you will receive 50 + 22 points = 72 points, and with *60 %* probability you will receive 50 – 20 points = 30 points. Participant A will receive 90 – 22 points = 68 points with *40 %* probability and his initial endowment of 90 points with *60 %* probability.

Example 2: You choose a transfer amount of 0 points. You will receive 50 points. Participant A will receive 90 points.

The course of action of part 1 is illustrated by the following graph:



If you have any questions, please raise your hand. An experimenter will come to your seat to answer your questions.

2.7.2 Translated Holt and Laury (2002) table

Decision	Option A	Option B
Decision 1	10 points	25 points with a probability of 10 % 0 points with a probability of 90 %
Decision 2	10 points	25 points with a probability of 20 % 0 points with a probability of 80 %
Decision 3	10 points	25 points with a probability of 30 % 0 points with a probability of 70 %
Decision 4	10 points	25 points with a probability of 40 % 0 points with a probability of 60 %
Decision 5	10 points	25 points with a probability of 50 % 0 points with a probability of 50 %
Decision 6	10 points	25 points with a probability of 60 % 0 points with a probability of 40 %
Decision 7	10 points	25 points with a probability of 70 % 0 points with a probability of 30 %
Decision 8	10 points	25 points with a probability of 80 % 0 points with a probability of 20 %
Decision 9	10 points	25 points with a probability of 90 % 0 points with a probability of 10 %
Decision 10	10 points	25 points with a probability of 100 % 0 points with a probability of 0 %

Participants made ten separate decisions whether they preferred option A to option B with the varying associated probabilities displayed above. One decision was chosen randomly (all with equal probability) and paid at the end of the experiment.

We classify the 51 subjects who prefer option A to option B in decisions 1 till 4 and option B to option A otherwise as risk neutral. The 16 subjects switching before decision number 5 are categorized as risk seeking. We have 88 risk averse subjects who switch from preferring option A to option B only at a decision number higher than 5. Three subjects behave irrationally (or are humble): they always prefer option A.

Chapter 3

Does parental employment affect children's educational attainment? Evidence from Germany

3.1 Introduction

Over the last decades, female labor market participation rates and especially those of mothers with young children have increased tremendously in many countries. In the US, 47 % of mothers with children below age 6 worked in 1975. By 2006, this share had increased to 71 % (Chao and Rones, 2007, Table 7). In Germany, 35 % of mothers with children below age 6 worked in 1974, but 52 % in 2004. In contrast, labor market participation rates of German fathers have remained very stable at about 88 %.[42]

Precise knowledge about how parental employment affects children's long-term outcomes such as educational attainments or labor market success is crucial for the evaluation of many policy programs. For example, US welfare reforms in the 1990s pushed welfare recipients and in particular welfare dependent single mothers to find employment (compare Blank, 2002). Reforms were motivated by the belief that parental work is the best way out of poverty for parents and children. If, however, having working

[42] Figures stem from an inquiry at the Federal Statistical Office of Germany. The 35 % in 1974 refer to West Germany only.

parents hurts the educational and labor market prospects of children such reforms may be counterproductive in the long run. To give another example, the current German government's decision to substantially expand and subsidize day care facilities for children below age 3 has lead to emotional and controversial debates in the German public. Opponents of day care expansion consider full-time parental child care to be decisive for children's cognitive and emotional development. Proponents argue that parent-child interactions can be substituted by high quality non-parental child care and that increases in family income may also benefit children.

This paper is the first to use a large German household panel data set, the German Socioeconomic Panel (GSOEP), to analyze whether parental employment hurts or benefits children's educational attainments. Our measure of educational attainment and dependent variable is attendance or completion of high secondary school track (so called Gymnasium) which is the only track that provides direct access to university. We separately analyze two effects of parental employment: first, the effect on income that may influence child-related investments, i.e. we control for total household income. Second, we use three different measures of parental time inputs in raising their children to capture the "time effect" of parental employment: weekly hours worked, the number of years in which parents work full-time, part-time or not at all and weekly hours that parents spent on child care when children are aged 0-3.

We explicitly approach potential endogeneity problems. First, to take selection of parents in the labor market into account we estimate a model on sibling differences that controls for all unobserved time-invariant parent and household characteristics. Second, we address the potential reverse causality problem, i.e. the fact that parents' decisions to work may be influenced by their child's ability which in turn affects educational attainment. We focus on parental employment when children are aged 0-3 such that a child's ability is not yet fully revealed, exclude disabled children from the analysis and use parents' years of education as a proxy for their child's ability.

We do not find any evidence that parental employment hurts children's educational attainment. Controlling for household income we can statistically rule out that having a mother who works one hour more per week lowers the probability of high secondary track attendance by more than 0.1 percentage points. Actually, all coefficients of maternal employment are positive, but not significant at conventional levels (though at a 9 to 11 % significance level). Coefficients of fathers' employment and parental time spent on child care are precisely estimated, but too small to be significant. Testing for equality of mother's and father's time input coefficients, we cannot reject that parents' time inputs are substitutes.

Table 3.1 reviews results from previous economic studies that investigate the relationship between parental employment and children's educational attainment. In sum, evidence is very inconclusive: some studies predict a negative effect of parental employment on children's educational attainment, some a positive one and the remainder insignificant effects or effects that differ by subsamples such as sex or race of the child. Table 3.1 also reveals some characterizing features of existing studies. First and most importantly, except for Ermisch and Francesconi (2002) the studies listed in Table 3.1 ignore problems that arise due to omitted variables such as a child's ability or selection of parents into the labor market. In contrast, this paper addresses the conditions under which we obtain consistent estimates explicitly and estimates a model on sibling differences to control for unobserved parent and household characteristics. Second, only two studies (Ermisch and Francesconi, 2002 and O'Brien and Jones, 1999) report estimates on the effect of father's employment. Our paper estimates the effects of parental employment separately for mothers and fathers and as Ermisch and Francesconi (2002) also the joint effect of e.g. hours worked. Third, all studies use US or British data. Since the institutional environment (child care facilities, maternity leave policies, etc.) and the attitudes towards working mothers differ substantially across countries, evidence from Anglo-Saxon countries might not be transferable to other Western countries. Our study adds evidence from Germany to the existing literature. Last, all studies in Table 3.1 use indirect measures of parental time inputs such as the type of parental employment (full-time, part-time or none) or years worked.[43] An advantage of the GSOEP data set is that it contains very detailed information on

[43] Haveman, Wolfe and Spaulding (1991) use *estimated* parental time spent on child care as explanatory variable. They do not have information on time spent on child care in their original data but construct it from a second data set. Using the second data set they regress child care time on explanatory variables that their original data and the second data set have in common and then apply coefficients to their original data set to construct estimates of child care time.

Table 3.1: Related literature

study	data source, country	outcome	estimation method	effect of parental employment*
Ermisch and Francesconi (2002)	British Household Panel Survey, UK	highest educational qualification (A level or more)	logit, linear probability models, sibling differences model	mother works part-time: level estimates: (-) ns sibling difference est.: (-) 10 mother works full-time: level estimates: (-) ns sibling difference est.: (-) 5 father works: level estimates: (+) 5 sibling difference est.: (-) ns
Graham, Beller and Hernandez (1994)	Current Population Survey, US	years of schooling at ages 16-20	2SLS, first stage: IV for child support	mother worked outside home: (+) 1
Haveman, Wolfe and Spaulding (1991)	Panel Study of Income Dynamics, US	high school graduation	probit model	years mother worked: (+) 1
Hill and Duncan (1987)	Panel Study of Income Dynamics, US	years of schooling at ages 27-29	OLS, gender specific	mother's work hours: for men: (-) 5 for women: (-) ns
Kiernan (1996)	National Child Development Study, UK	no degree	descriptive statistics, logit model	mother's non-employment: for men: no effect for women: (+) 1
Krein and Beller (1988)	National Longitudinal Surveys, US	years of schooling at age 26	OLS, gender and race specific	mother ever worked outside home at least 6 months at ages 0-18: white men: (-) 1 white women: (-) ns black men: (+) ns black women: (-) ns
O'Brien and Jones (1999)	survey and time-use diaries in 6 schools in East London, UK	highest / lowest national test scores	logit model	low educational outcome: father works full time and mother full time: (-) ns mother part-time: (-) 5 high educational outcome: father works full-time and mother full time: (+) ns mother part-time: (+) 10

* (-) indicates a negative sign, (+) a positive sign
ns: not significant; 1, 5, 10: significant at a 1, 5, 10 percent significance level

the time parents spent on child care. Besides the commonly used indirect measures we use the hours parents spent on child care on a typical weekday when children are aged 0-3 as a direct measure of parental time inputs in raising their children.

Haveman and Wolfe (1995) review studies on the effects of parental employment on a broad range of children's outcomes such as high school graduation, years of schooling, out-of-wedlock fertility or adult earnings. All reviewed studies use US data and do not address endogeneity problems. Ermisch and Francesconi (2005) survey the more recent literature on parental employment and children's well-being covering studies that use data from countries different from the US, mainly from UK.

Using German administrative data Dustmann and Schönberg (2007) analyze the effect of three extensions in maternity leave coverage on children's later attendance of high secondary track and wages. They compare cohorts of children born shortly before and after the reforms. Although reforms induced women to delay their return to work, the authors do not find that an expansion in maternity leave legislation improves child outcomes. By exploiting unexpected changes in legislation the authors can nicely infer causal effects at the cohort level. We consider our approach complementary to theirs: using individual level instead of cohort data we can evaluate the importance of numerous individual and family characteristics for child outcomes.

A couple of papers use GSOEP data to explain high secondary track attendance, but none analyzes the impact of parental employment. Büchel and Duncan (1998) explore the role of parents' social activities (e.g. socializing with friends, attending cultural events, doing volunteer work), Francesconi, Jenkins and Siedler (2006) the impact of growing up in a family headed by a single mother and Tamm (2007) investigates the effect of parental income on high secondary track attendance.

The structure of the paper is as follows: section 2 offers basic information on the German school system, section 3 provides a brief overview on the GSOEP data set. Economic framework and estimation methods are discussed in section 4. In section 5, we present results from level and sibling difference regressions as well as non-parametric Kernel density estimates. Section 6 concludes.

3.2 Institutional background: the German school system

In Germany, all students jointly go to elementary school for at least four years. After elementary school, usually at age 10, students proceed to secondary education. The secondary education system is organized in three main tracks: Least academic and most vocational general secondary school track (Hauptschule, grades 5 to 9) provides basic secondary education and prepares for an apprenticeship in a blue collar job. Intermediate secondary school track (Realschule, grades 5 to 10 or 11) is usually followed by an apprenticeship in a white collar job. Only students of the most academic track, the high secondary school track (Gymnasium, grades 5 to 12 or 13), obtain a final degree that provides access to university.

Education is regulated by the states (Bundesländer). In all states, track choice after elementary school is influenced by a recommendation of elementary school teachers that is mainly based on performance and the decision of parents. To which extent parents can influence their child's school track differs substantially across states. In some states the tracking decision is delayed from fourth to sixth grade, and all students jointly go to Förderschule in fifth and sixth grade. Furthermore, some states have a comprehensive school type (Gesamtschule) that comprises all three tracks. All states have schools for children with special needs due to physical or mental disabilities (Sonderschule). Finally, there are very few so-called Waldorf schools that are private and follow a special pedagogy. Still, in our data, about 88 % of students are part of the standard three track system: 20 % attend general secondary school track, 34 % intermediate and high secondary school track each. In all states, secondary education is compulsory up to grade 9 and provided free of charge.

Changing secondary school track after initial choice is possible, but relatively rare. Using GSOEP data on West Germans born between 1970 and 1984, Tamm (2007) compares secondary school tracks attended at age 14 with the highest secondary school degree obtained at age 21. He finds that between 60 % and 70 % of students obtain the degree of the secondary school track they attended at age 14. There is some upward mobility: 21 % (5 %) of those attending intermediate (general) secondary school at age 14 manage to obtain a degree which provides complete or restricted access to university. In each school track roughly 10 % drop out without any degree. Schnepf (2002) provides further evidence on low rates of track changing.

The dependent variable of our analysis is secondary school track, more precisely whether a child attends high secondary track or does not. Secondary school track is an important determinant of labor market outcomes later in life. Using GSOEP data, Dustmann (2004) shows that having successfully attended the high (intermediate) instead of the general secondary school track increases the wage at labor market entry by 29.3 % for men and 37.7 % for women (15.9 % for men and 26.7 % for women, respectively). This holds true even when controlling for post-secondary education that is strongly influenced by secondary school track. The wage premium increases to far more than 50 % for a high instead of general secondary education degree when post-secondary education is not controlled for.

3.3 Data

We use data from the German Socioeconomic Panel (GSOEP). The GSOEP is a representative panel study of German households that covers the years 1984 until 2006. In addition to household level information, individual information is available. Data cover a wide range of topics such as individual attitudes and health status, job characteristics, unemployment and income, family characteristics and living conditions. For children up to age 15, personal information is provided by the head of the household. We use subsamples A to D, i.e. data on households living in East and West Germany[44] irrespective of their nationality. Haisken-DeNew and Frick (2003) provide a detailed description of the GSOEP.

Our dependent variable is binary and indicates whether a child attends high secondary school track or does not, i.e. attends general or intermediate secondary education. Hence, it focuses on whether children will obtain access to university after

[44] Sampling of East German households started in 1990.

finishing school or do an apprenticeship as both general and intermediate secondary track students usually do.[45] We use the latest available information on track choice to minimize inaccurateness caused by later changing of tracks. Children attending other types of schools (such as Gesamtschule, Förderschule, Waldorfschule or Sonderschule) are excluded from our analysis.

Parents' time inputs are the primary variables of interest. We use three alternative variables to check the sensitivity of our results: (i) weekly hours worked, (ii) the number of years in which parents have a full-time, part-time or no job, and (iii) hours spent on child care on an average weekday. While average hours spent on child care is the most direct measure, it is also the most subjective one. Some parents claim to devote 24 hours per day to child care, others, who also stay at home, state much lower numbers. In contrast, type of employment and hours worked are more objective measures. They do not capture time inputs directly, but are strongly negatively correlated with hours spent on child care: for fathers, the correlation coefficient ρ between hours worked and time spent on child care is -0.34 and significantly different from zero ($p < 0.001$), for mothers, $\rho = -0.32$ with $p < 0.001$.[46] For the largest part of our analysis we use averages of one of these three variables over a child's first three years. There are two reasons for focusing on the first three years. First, this is the period that is most debated in public - for example, up to now public child care facilities in (West) Germany have nearly exclusively been available for children from age three onwards. Second, as will become clear in the next paragraph, our identifying assumption is that parents do not know their child's ability as long as their child is sufficiently small and, thus, cannot condition their employment decision on their child's ability. This assumption is more plausible the younger a child is.

A list of all explanatory variables, their means and sample sizes is displayed in Ta-

[45] Reducing the three track system to a binary dependent variable makes our results better comparable to those of the related literature, see for example Puhani and Weber (2007), Büchel and Duncan (1998) and Francesconi, Jenskin and Siedler (2006). Furthermore, results of a model with a binary dependent variable are easier to interpret than those of an ordered logit model. While it would in principle be possible to estimate an ordered logit model on sibling differences, this would require the additional assumption that the difference between general and intermediate secondary track is the same as the difference between intermediate and high secondary track.

[46] Similarly, the correlation coefficients between working full-time and average hours spent on child care per day are $\rho = -0.32$ for fathers and $\rho = -0.28$ for mothers, both with $p < 0.001$.

ble 3.2. As robustness checks, Tables 3.8 and 3.9 in the appendix present the complete results of our main specification when using data on non-foreign West Germans (subsample A) only (Table 8, column (2)), when using school track information at age 14 (instead of latest available information) as dependent variable (Table 8, column (3)) or when using parents' year-specific time use and employment information (Table 9). For all robustness checks, magnitudes of coefficients and significance levels do not change substantially compared to the baseline specification.

Table 3.2: Summary statistics

variable	general sample	siblings sample
information on the child		
attends high secondary school track	0.356	0.329
male	0.500	0.507
year of birth	1989.614	1989.835
firstborn child	0.484	0.402
born from January till June	0.508	0.513
information on the household		
total monthly net equivalent income*,**	0.917	0.924
non-German household	0.238	0.238
information on the mother		
age at birth <=21	0.117	0.127
age at birth 22-35	0.829	0.838
age at birth >36	0.053	0.035
years of education	11.486	11.549
weekly hours worked*	7.743	5.196
time spent on child care per weekday*	8.839	9.454
not working (number of years)	2.189	2.338
part-time job (number of years)	0.477	0.451
full-time job (number of years)	0.334	0.211
information on the father		
age at birth <=21	0.028	0.025
age at birth 22-35	0.829	0.864
age at birth >36	0.143	0.111
years of education	11.949	11.993
weekly hours worked*	40.363	40.182
time spent on child care per weekday*	2.235	2.288
not working (number of years)	0.196	0.198
part-time job (number of years)	0.028	0.048
full-time job (number of years)	2.775	2.754
N***	1047	550

* average at ages 0-3 of child
** in 1000 Euros
*** deviant number of observations for time spent on child care (N=1032 and 537) and for type of employment (N=962 and 521)

3.4 Economic framework, identification and estimation

Why should parental employment affect children's educational attainments? The very stylized and simplified static framework underlying our empirical analysis assumes that children's educational attainment s_i is a function of parents' time inputs, t_i, and goods and services inputs, x_i, and the child's ability, μ_i: $s_i = f(t_i, x_i, \mu_i)$ where all three first partial derivatives are positive. Both time and good inputs are influenced by parents' employment decisions. On the one hand, we expect parents who work to spend less time with their children (e.g. to play with them or to educate them) which results in a negative "time effect". On the other hand, we expect a positive "input effect". The more parents work the higher is the family income. Due to the income effect normal good inputs such as the number of books and toys at home or extra lessons in the afternoon will increase (if parents are altruistic at least to some degree). In our regressions, we will use family income as (the best available) proxy for goods and services inputs.

Our framework is most closely related to Leibowitz (1974) who assumes that family income has an additional direct impact on the schooling level. A similar relationship can also be derived from a production function framework that draws an analogy between the knowledge acquisition process of an individual and the production process in a firm (see, for example, Todd and Wolpin, 2003). The theory of family behavior (see Becker and Tomes, 1986, 1979 and Solon, 1999 for a simplified version) assumes that parents' intertemporal utility depends on their own consumption and on children's outcomes that are increasing in monetary investments in children. Consequently, parents invest part of their earnings in their children to maximize their own utility subject to a budget constraint. This gives the input effect. The time effect could be obtained by adding a time constraint and time investments to the model.

To begin with we estimate the following logistic regression model for a child i from family j:

$$(1)\ \Pr(high_{ij} = 1 | \underline{X_{ij}}, \underline{X_j}) = F(\beta_0 + \underline{\beta_1} \underline{X_{ij}} + \underline{\beta_2} \underline{X_j})$$

where $F(z) = \frac{\exp(z)}{1+\exp(z)}$ is the standard logistic distribution.

$High_{ij}$ is a binary variable that equals one if a child attends or has already finished high secondary track and zero otherwise. $\underline{X_{ij}}$ is a vector of characteristics that differ for different children of one family. It contains (i) child characteristics, namely a child's year of birth (normalized by subtracting 1984, the first year observed in our data) and binary indicators of a child's sex, whether a child is the firstborn child and whether a child is born between January and June[47], (ii) total net equivalent income of the household averaged over the ages 0-3 of the child and (iii) time varying parent characteristics: separate indicators for whether father and mother were younger than 22 or older than 36 when the child was born as well as information on mother's and father's employment or time spent on child care at ages 0-3 of child i.[48] $\underline{X_j}$ is a vector of characteristics that are shared by different siblings of one family j. It encompasses (i) household characteristics, here whether the household is classified as foreign (subsamples B and D in the GSOEP data) and (ii) time invariant characteristics of parents (father's and mother's total years of education measured as schooling plus apprenticeship plus university studies) and (iii) a vector of state dummies. β_0 is a constant and $\underline{\beta_k}, k = 1, 2$ are vectors of unknown parameters.

The coefficients of interest are those of parental employment or time spent on child care. Since we control for household income, they measure the time effect of parents' employment on a child's track choice.

To identify the true underlying coefficients we need to address potential endogeneity problems due to omitted variables. First, since a child's ability is unobserved coefficients of explanatory variables that are correlated with ability may be inconsistently estimated. In particular, this might be the case for the effect of parental employment if parents condition their employment on their child's ability (reverse causality). To give

[47] In Germany, children born between January and June (July and December) usually enter school in autumn of the year in which they become six (seven) years old. Puhani and Weber (2007) show that children who enter school at an older age because they are born between July and December perform better at school and have a higher probability of attending high secondary track.

[48] Averages over household income and time input information are taken over the years in which the information is available, i.e. for some observations we just observe information in one or two out of three years. For both household income and time input variables results are robust if we use only those observations for which information is available for all three years.

an extreme example, parents with disabled children might not work at all. We exclude children attending Sonderschule, i.e. disabled children or children with very low ability from our analysis. Apart from these extreme cases, our identification strategy assumes that parents do not know their child's ability as long as their child is sufficiently young and, thus, cannot condition their employment decision on their child's ability. The idea is that information revelation takes time and the amount of feedback increases only as a child grows.[49] To make this argument plausible we exclusively use parents' labor market participation when children are aged 0-3. As an additional safeguard, we include parents' education as a proxy variable for their child's ability. Here, we exploit that parents' education is correlated with their own ability which in turn is partially inherited by their children.

Second, selection of parents in the labor market is a potential problem. Imagine that parents with unobserved characteristics u_j that are either especially supportive or detrimental to raising children systematically decide (not) to work. If this were the case, the coefficients of parents' employment would not only capture the time effect that we would like to measure, but also the effect of parents' unobserved characteristics on the child's educational attainment. To control for all time-invariant unobserved parent characteristics we estimate a model on sibling differences (compare Ermisch and Francesconi, 2002 for a similar application). Identification rests on the assumption that parents' relevant unobserved characteristics summarized in u_j (e.g. the quality of parent-child interactions) do not vary for different siblings.

To estimate a model on sibling differences we drop all observations on children without siblings from the sample. We sort all children of a family (who are born between 1984 and 1996 and for whom complete information is available) by age and build pairs of siblings. A pair consists of two adjacent siblings such that we get $n-1$ differences for n siblings. To be able to interpret our results in terms of probability of high secondary track attendance we need to estimate a binary model. For this purpose we construct sorted differences: first, we subtract values of all variables that belong to the older sibling from the values of the corresponding variables of the younger

[49] Our argument is similar to Rosenzweig and Wolpin (1995) who argue that parents update their beliefs about their children's endowments as time passes.

sibling. This results in a new differenced dependent variable that takes values -1, 0 or 1. Second, to be back in a binary framework we multiply all (dependent and explanatory) differenced variables of a sibling pair with -1 if the differenced dependent variable is originally -1 ("sorting"). Equation (2) illustrates this procedure for the oldest two siblings of a family in a linear probability model:

$$(2)\ \Pr\{\mathbf{1}(high_{2j} - high_{1j}) = 1 | \mathbf{1}(\underline{X_{2j}} - \underline{X_{1j}})\} = \widetilde{\beta_0} + \widetilde{\underline{\beta_1}}(\mathbf{1}(\underline{X_{2j}} - \underline{X_{1j}}))$$

$$\text{with } \mathbf{1} = \begin{cases} 1 & if \quad (high_{2j} - high_{1j}) \in \{0,1\} \\ -1 & if \quad (high_{2j} - high_{1j}) = -1 \end{cases}.$$

The numbers 1 and 2 index the first and second born sibling respectively. By construction, all explanatory variables that have the same value for siblings, i.e. u_j and $\underline{X_j}$, are dropped in a sibling difference estimation. Differences in parents' age at birth are collinear with the difference in years of birth of two siblings and are also dropped. While the original constant term β_0 disappears due to differencing, we include a new constant term $\widetilde{\beta_0}$ to account for the effect of being sorted first (compare Ermisch and Francesconi, 2002 and Ashenfelter and Rouse, 1998). The new constant term $\widetilde{\beta_0}$ arises from differencing a dummy variable that is equal to one if a sibling is sorted first and zero otherwise. $\widetilde{\beta_0}$ captures that due to sorting the sibling sorted first in the difference has a higher probability of attending high secondary track. The interpretation of the sign, but not the level of the coefficients in the difference model is the same as in the level model. For example, imagine that we had estimated a significantly positive coefficient β_k for the effect of the difference in maternal weekly hours worked in a linear probability model. Then β_k would imply that having a mother who works one hour more when sibling 2 (sorted first) is small than when sibling 1 is small increases the probability that sibling 2 attends high secondary track while sibling 1 does not by $\beta_k \times 100$ percentage points. Hence, a positive (negative) sign still stands for a positive (negative) effect.

In contrast to Ermisch and Francesconi (2002), we estimate a linear probability model on the differences and not a logit (or probit) model. The reason is that with a logit or probit model the assumption that the error term has a standard logistic or standard normal distribution will either be true for the original level or the difference model.[50] The main disadvantage of a linear probability model is that it may predict probabilities larger than unity or smaller than zero for extreme values of explanatory variables. Since we are interested in the average marginal effects this is not a major problem. Furthermore, we will check how close the estimates of the average marginal effects in a linear probability model are to those obtained in a probit model.

A more commonly used alternative to a sibling difference model is a household fixed effect model (conditional logit model). In our application, a conditional logit model uses only those observations on sibling pairs in which one sibling attends high secondary track and the other one does not. It estimates coefficients by comparing sibling pairs in which the older sibling attends high secondary track to sibling pairs in which the younger sibling attends high secondary track. Thus, identification of the effect of parental employment on children's educational attainment stems from observations in which both educational outcome and parental employment differ across siblings. In contrast, our sibling difference model also uses observations from families in which all children go to the same secondary track and estimates coefficients by comparing siblings pairs in which both siblings do or do not attend high secondary track to sibling pairs in which only one sibling attends high secondary track. Consequently, the identification of the effect of parental employment on children's educational attainment stems from all observations in which parental employment differs across siblings. We prefer estimating a sibling difference model to estimating a household fixed effect model because the former uses more observations which allows estimating coefficients more precisely.

[50] The variance of the difference of two random variables is equal to the sum of the two variances minus the covariance. Thus, the variance of the difference of two random variables with a standard logistic (normal) distribution only equals the variance of the standard logistic (normal) distribution if the covariance incidentally coincides with the standardized variance.

3.5 Results

3.5.1 Estimation on levels

We will first present results from a logit estimation (Table 3.3) that does not address endogeneity problems caused by unobserved parent characteristics. The results are still a useful benchmark for comparison with other studies that use similar specifications. Additionally, results provide some information on the coefficients of variables that are constant for siblings and thus will drop out when estimating a sibling difference model.

While the coefficient of mother's average hours worked is not significant, the coefficient of father's average hours worked is weakly significant (p=0.080) and positive. Setting all control variables to their mean the predicted marginal effect if the father would work one hour more per week in every year is a 0.5 % increase in the probability that his child attends high secondary track. Furthermore, male children are predicted to attend high secondary track less often than female children. Firstborn children are more likely, children born between January and June are marginally less likely to attend high secondary track. In Puhani and Weber (2007) children who enter school at an older age because they are born between July and December have a 12 % higher probability of attending high secondary track. If control variables are set to their mean, our results predict a 9 % higher probability of attending high secondary track. Having a father who is at most 21 in the year of birth is predicted to have an adverse effect on the child's educational attainment, having a relatively old mother seems to be supportive. Both coefficients are likely to suffer from endogeneity problems and hence might reflect unobserved parents' characteristics that are correlated with the included age intervals. The coefficients of parents' total years of education are highly significant and positive. We use state dummies to control for state specific differences in shares of students attending high secondary track.[51]

[51] Results reported in Table 3.3 are very robust to using mother's and father's age and age squared instead of age intervals, including year dummies instead of imposing a linear time trend or to including dummies for the number of siblings which we do not do because it reduces the number of observations by about 10 %. Furthermore, using an ordered logit specification with a dependent variable that takes the values 1 to 3 for general, intermediate and high secondary track attendance produces estimates very similar to those reported in Table 3.3.

Table 3.3: Base specification: logit estimation on levels

binary dependent variable: child attends high secondary track

explanatory variables	coefficient	p-value
mother's weekly hours worked*	0.000	0.960
father's weekly hours worked*	0.019	0.080
male	-0.639	0.008
born before July	-0.380	0.095
firstborn child	0.635	0.003
year of birth - 1984	-0.006	0.889
age of mother at birth <=21	0.598	0.133
age of mother at birth >36	0.987	0.034
age of father at birth <=21	-3.822	0.000
age of father at birth >36	0.259	0.487
mother's total years of education	0.420	0.000
father's total years of education	0.348	0.000
household income*,**	0.086	0.927
(household income)^2 *,**	-0.080	0.785
non-German household	0.180	0.690
constant	-11.136	0.000
state dummies	yes	
N	1047	
Pseudo R^2	0.393	

* average at ages 0-3 of child
** total monthly net household equivalent income in 1000 Euros
comment: robust, clustered standard errors that allow observations to be correlated within a family

Table 3.4 presents the coefficients of parental time inputs. The upper part uses hours worked per week averaged over the ages 0-3 of a child as measure of time inputs, the middle part hours spent on child care on a typical weekday averaged over the ages 0-3 of a child. The lower part presents coefficients of type of employment, e.g. variables that indicate how many out of a child's first three years parents did work full-time, part-time or not at all. The omitted categories are the most common ones: working full time for fathers and not working at all for mothers. The table contains the coefficients from three different specifications. Specification A uses observations from families in which both parents are present (in our data) and information on both parents' time inputs, age and education is completely available. As most studies on the relationship between parental employment and children's educational attainment specification B does not include information on father's age, education and time inputs as explanatory variables. Hence, it additionally includes observations from families with single mothers and families with present fathers on whom information is incomplete which raises the number of observations by about 20 %. We add a dummy variable for absent fathers that turns out not to be significant. Specification C adds a dummy variable that takes a value of one if the father is present and zero otherwise and reports coefficients of father's age, education and time inputs when age, education and time inputs are interacted with this dummy. Otherwise explanatory variables in Table 3.4 are the same as in Table 3.3.

In the eight additional specifications, all coefficients of parental time inputs are not significant with one exception: the coefficient of mother's full time employment is negative and significant (p=0.022) in specification B.

3.5.2 Estimation on sibling differences

The sample

The siblings sample contains data on 550 siblings from 249 families. Table 3.2 compares means in the general and the siblings sample. The siblings sample is largely representative for the general sample. Differences in means usually occur only in the

Table 3.4: Further specifications: logit estimation on levels

binary dependent variable: child attends high secondary track

coefficients of average weekly hours worked

specification	(A)	(B)	(C)
mother	0.000	-0.008	0.006
	(0.960)	(0.327)	(0.487)
father	0.019	-	0.012
	(0.080)	-	(0.297)
N	1047	1280	1214
Pseudo R^2	0.393	0.305	0.354

coefficients of type of employment in years

specification	(A)	(B)	(C)
full time mother	-0.189	-0.253	-0.198
	(0.210)	(0.022)	(0.158)
part time mother	0.044	-0.044	0.088
	(0.795)	(0.739)	(0.568)
part time father	-0.422	-	-0.493
	(0.530)	-	(0.407)
non-working father	0.200	-	0.239
	(0.352)	-	(0.276)
N	962	1195	1118
Pseudo R^2	0.388	0.335	0.372

coefficients of average hours spent on child care per weekday

specification	(A)	(B)	(C)
mother	0.050	0.031	0.005
	(0.156)	(0.300)	(0.867)
father	-0.110	-	-0.096
	(0.092)	-	(0.157)
N	1032	1262	1199
Pseudo R^2	0.387	0.302	0.348

comments: robust, clustered standard errors that allow observations to be correlated within a family; p-values are reported in brackets
(A): uses only observations with complete information on both parents
(B): uses all observations with complete information on mother
(C): estimates coefficients of father's age, education and time inputs conditional on father being present

second position after the decimal point. Of course, the sibling sample contains fewer firstborn children (40 % instead of 48 %). On average, mothers in the siblings sample work 2.5 hours less per week and spend 0.6 additional hours per day on child care. Father's employment is very similar in both samples.

Kernel density estimates

To get a first impression whether differences in parental employment could be driving differences in siblings' educational attainment we estimate non-parametric Kernel densities. The solid line depicts sibling pairs who either both attend high secondary track or both do not. The dashed line stands for sibling pairs in which one sibling attends high secondary track, but the other one does not. Again, the sibling attending high secondary track is sorted first in the difference. Figure 3.1 (Figure 3.2) displays Kernel density estimates of the distributions of differences in average hours worked by mothers (fathers) when children were aged 0-3 for these two kinds of sibling pairs. If having a mother or father with longer working hours would reduce the attendance of high secondary track we would expect the dashed line to be first order stochastically dominated by the solid line.

Figure 3.1: Kernel density estimates, mother's hours worked

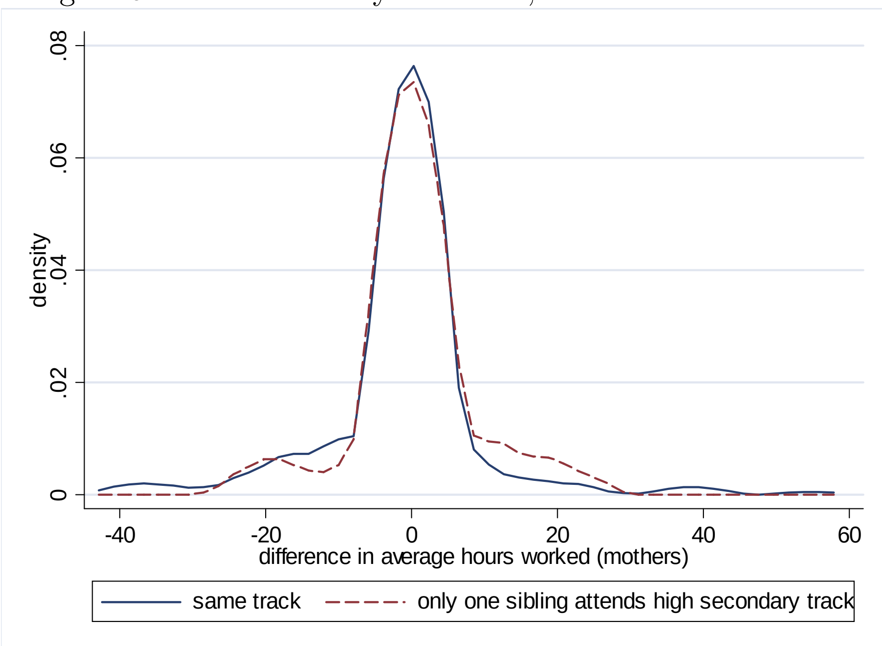
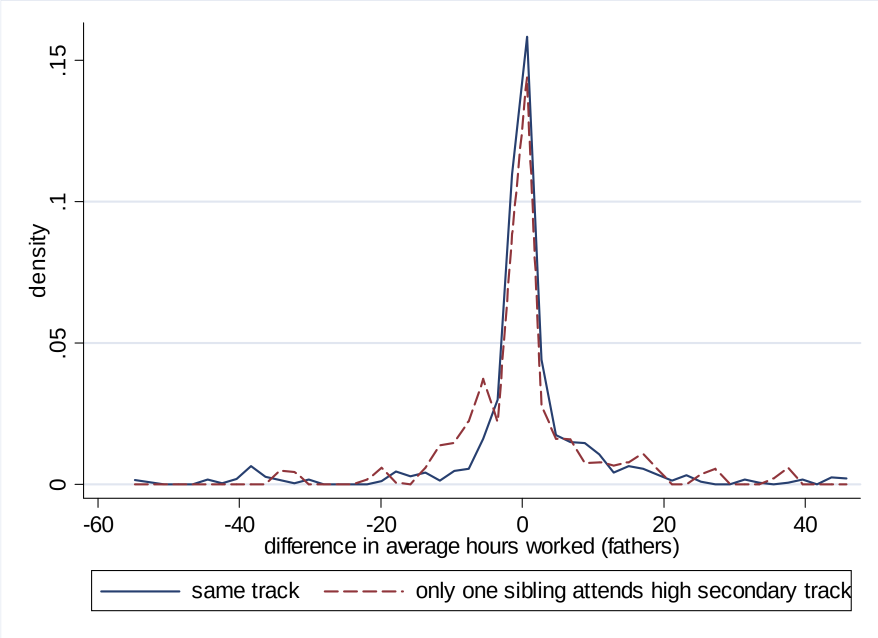


Figure 3.2: Kernel density estimates, father's hours worked



Eyeballing suggests that estimated densities are very similar. Non-parametric, two-sided Mann-Whitney and Kolmogorov-Smirnov tests on the original (not the displayed estimated) distributions confirm that distributions do not differ significantly: for mothers $p_{MW} = 0.394$ and $p_{KS} = 0.824$ and for fathers $p_{MW} = 0.740$ and $p_{KS} = 0.404$. At first sight, differences in parental employment patterns do not seem to be a driving force behind different levels of educational attainments.[52]

Multivariate analysis

To control for differences between siblings apart from parental employment we estimate a linear probability model on sibling differences to explain different educational outcomes. The estimation requires sufficient variation in both dependent and explanatory variables. In all specifications, we have about 20 % of sibling pairs in which just one sibling attends high secondary track. Table 3.5 and Figures 3.1-3.4 document substantial variation in mother's and father's average hours worked as well as in time spent on child care. By construction variation is largest in working hours that are measured

[52] Appendix 3.7.1 displays estimated Kernel densities for the average time parents spent on child care.

per week, followed by hours spent on child care measured at a daily level. Variation is smallest for type of employment that is measured in years such that differences can at most range from -3 to 3. For this reason we will provide estimates on sibling differences only for the former two measures of parental time inputs (in contrast to Ermisch and Francesconi (2002) who use the difference in type of employment).

For our baseline specification, Table 3.6 compares the marginal effects predicted by an estimation on sibling differences using the linear probability or the probit model. The marginal effects - especially the significant ones - are very similar. Thus, the usual drawbacks of estimating a linear probability model are not a major concern in our application and for the reasons outlined in section 3.4 we prefer the linear probability model.

Results in Tables 3.6 and 3.7 show that differences in father's employment do not contribute significantly to explaining differences in educational attainment. The coefficient of differences in mother's employment is positive and just not significant (in Table 3.7, p=0.105 in specification (A) and (C) and p=0.085 in specification (B)). The precision of our baseline estimate (specification (A)) implies that we can statistically rule out that having a mother who works one hour more per week (when the sibling sorted first was young than when the second sibling was young) lowers the probability that the sibling who is sorted first attends high secondary track (while the second sibling does not) by more than 0.1 percentage points. Similarly, the alternative specifications in Table 3.7 show that average time spent on child care does not influence attendance of high secondary track significantly.[53] As in the level estimation our sibling difference estimates in Table 3.6 predict children who are born between January and June to be less likely and firstborn children to be more likely to attend high secondary track. The advantage of firstborn children even increases with each year they are apart from the second born sibling. In contrast to the level estimation, the sex of a child is no longer significant.

The size of coefficients in the level and the difference estimation is not directly comparable since they refer to characteristics measured in levels or differences between

[53]Results in Table 3.7 are robust to adding a squared term for mother's and father's time inputs.

Table 3.5: Variation in key explanatory variables

differenced variable	mean	standard deviation	zeros (%)	min	max	N
mother's hours worked	-.822	9.960	57.48	-40	55	301
father's hours worked	.106	10.732	17.28	-54	45	301
mother's hours spent on child care	.525	3.497	16.61	-10	19	295
father's hours spent on child care	-.167	2.008	17.63	-9	10	295
mother's full time employment	-.076	.601	86.59	-3	3	276
mother's part time employment	.007	.772	70.65	-3	3	276
non-working mother	.069	.906	65.22	-3	3	276
father's full time employment	.025	.581	87.68	-3	3	276
father's part time employment	-.007	.256	97.46	-3	2	276
non-working father	-.007	.490	90.58	-3	3	276

Table 3.6: Linear probability and probit model on sibling differences

dependent variable: sibling difference in high secondary track attendance

model	linear probability			probit	
differenced variables	coefficient	p-value	95 % confidence interval	marginal effects*	p-value
mother's hours worked**	0.005	0.105	[-0.001, 0.012]	0.005	0.073
father's hours worked**	0.004	0.529	[-0.008, 0.015]	0.002	0.589
male	0.009	0.715	[-0.040, 0.058]	0.016	0.488
born before July	-0.097	0.032	[-0.185, -0.009]	-0.093	0.008
firstborn child	0.151	0.003	[0.050, 0.251]	0.089	0.015
year of birth - 1984	-0.042	0.005	[-0.071, -0.012]	-0.029	0.030
household income***	0.019	0.961	[-0.740, 0.777]	0.032	0.928
(household income)^2 ***	0.026	0.837	[-0.221, 0.273]	0.003	0.977
constant	0.317	0.000	[0.228, 0.406]	-0.063	0.000
N	301			301	
R^2	0.240			0.251	

* all other explanatory variables are evaluated at their mean
** average per week at ages 0-3 of child
*** total monthly net household equivalent income in 1000 Euros, average at ages 0-3 of child
comment: robust, clustered standard errors that allow observations to be correlated within a family

Table 3.7: Further specifications: linear probability model on sibling differences

dep. var.: sibling difference in high secondary track attendance

coefficients of difference in hours worked, average per week

specification	(A)	(B)	(C)
mother	0.005	0.006	0.005
	(0.105)	(0.085)	(0.105)
father	0.004	-	0.003
	(0.529)	-	(0.582)
N	301	372	345
R^2	0.240	0.307	0.321

coefficients of difference in time spent on child care, average per weekday

specification	(A)	(B)	(C)
mother	-0.003	-0.003	-0.002
	(0.747)	(0.688)	(0.772)
father	-0.016	-	-0.012
	(0.361)	-	(0.478)
N	295	361	338
R^2	0.239	0.293	0.314

comments: robust, clustered standard errors that allow observations to be correlated within a family; p-values are reported in brackets
(A): uses only observations with complete information on both parents
(B): uses all observations with complete information on mother
(C): estimates coefficients of father's age, education and time inputs conditional on father being present

siblings, respectively. Still, the interpretation of coefficients' signs (i.e. whether we observe a negative or positive effect) is comparable. The significance levels and, thus, implications from the level and difference estimation differ markedly. Table 3.2 documents that differences are not caused by different sample characteristics. This suggests that controlling for unobserved parent characteristics affects results and should become the standard in the literature on the effects of parental employment on children's educational attainment.

Furthermore, mother's and father's time inputs do not seem to influence children's educational outcomes in different ways: we can reject the hypothesis that the coefficients of mother's and father's employment and time spent on child care differ significantly (F-tests for specifications (A) in Table 3.7 yield F=0.09, p=0.763 for hours worked and F=0.33 and p=0.565 for time spent on child care). Table 3.8, Columns (4) and (5) in Appendix 3.7.2 display estimates for parents' joint working hours and joint time spent on child care. Since the coefficients of joint time inputs are not significant, they confirm our previous results.

Additionally, we check whether the effect of parental employment differs at different ages of the child, in our case at age 1, 2 and 3 (see Appendix 3.7.2, Table 3.9). Some studies that focus on children's short term outcomes have found that maternal employment during the first year of a child is especially detrimental. For example, Ruhm's (2000) and Waldfogel, Han and Brooks-Gunn's (2002) results imply that maternal employment during the first year of a child reduces math, reading and verbal achievement test scores at the ages 3-8 substantially. Our results on long-term outcomes are not consistent with this finding. In contrast, the coefficient of maternal working hours during the first year is marginally significant (p=0.082) and positive. F-tests for equality of coefficients document that each parent's coefficients do not differ across the ages 0-3 of a child. This justifies our approach of averaging time input information over the first three years.[54]

[54] For the three coefficients of mother's (father's) hours worked F=1.02 and p=0.364 (F=0.41 and p=0.666), for the three coefficients of mother's (father's) time spent on child care F=0.54 and p=0.586 (F=0.06 and p=0.939). Furthermore, in both specifications all six parents' coefficients do not differ significantly (F=0.58 and p=0.713 for hours worked and F=0.42 and p=0.832 for time spent on child care, respectively).

3.6 Concluding remarks

This paper has analyzed whether parental employment affects children's educational attainment. We have explicitly addressed potential endogeneity problems: to control for unobserved parent characteristics we have used estimates on sibling differences. To avoid inconsistent estimates due to reverse causality we have dropped disabled children from the analysis, have focused exclusively on parental employment when children are young (aged 0-3) such that signals about ability are still scarce and have included parent's education as a proxy variable.

Our measures of parental time inputs exclusively capture quantity, not quality - though quality is controlled for in the sibling difference estimates to the extent quality of parent-child interactions does not differ for different siblings. Due to lack of data, we have not controlled for non-parental time and good inputs and have ignored potentially important differences between different kinds of non-parental child care (such as attendance of Kindergarten, child care by relatives or nannies). These are important issues left for future research. Still, it is often argued that parental employment patterns per se shape a child's environment and outcomes. This is what we have tested for.

In sum, our results do not support worries that parental employment is detrimental for children's educational attainment. The core of our analysis are the estimates on sibling differences that use average weekly working hours when the child is aged 0-3 to measure parental time inputs: given their precision, we can statistically rule out that having a mother who works one hour more per week lowers the probability of high secondary track attendance by more than 0.1 percentage points, an economically negligible number. Actually, all coefficients of maternal employment are positive but not significant at conventional significance levels (though at an 9 to 11 % level). The corresponding coefficients of paternal employment and estimates using parental time spent on child care instead of working hours are not significant. Taken together, our results imply that it is not parental employment or quantity of parent-child interactions that are decisive for children's educational attainments, but, for example, birth order within a family, age relative to classmates or parental characteristics.

With respect to the current debate about the expansion of day care facilities in Germany our results do clearly not support worries that a more comprehensive child care infrastructure will hurt children's future prospects by raising maternal employment. Of course, our estimates are based on data from the past. To some extent, the current reforms will lead to changes in the institutional environment and perhaps also society's attitudes towards working mothers that may affect the interplay between parental employment and child outcomes.

3.7 Appendix

3.7.1 Kernel density estimates for time spent on child care

Figure 3.3 (Figure 3.4) displays Kernel density estimates of the distributions of differences in average hours spent on child care by mothers (fathers). The dashed line depicts sibling pairs in which one sibling attends high secondary track and the other one does not. The solid line marks siblings who either both attend high secondary track or both do not.

Figure 3.3: Kernel density estimates, mother's time spent on child care

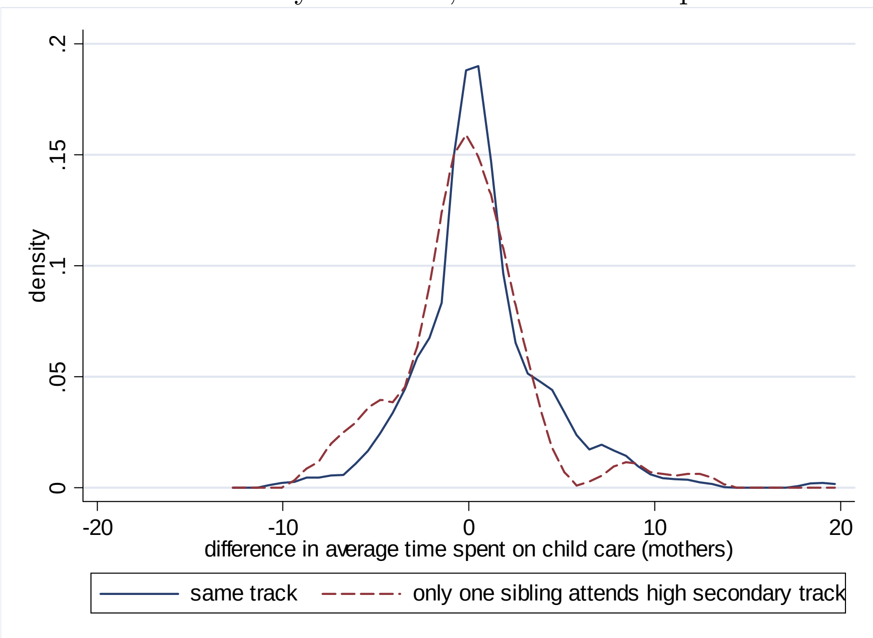
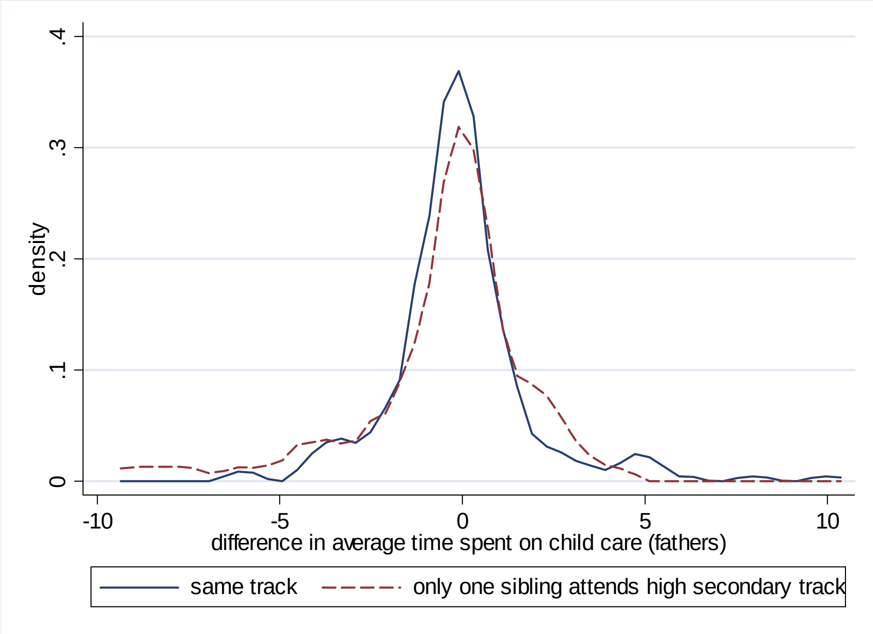


Figure 3.4: Kernel density estimates, father's time spent on child care



Non-parametric, two-sided Mann-Whitney and Kolmogorov-Smirnov tests on the original distributions confirm that distributions do not differ significantly for fathers, $p_{MW} = 0.953$ and $p_{KS} = 0.884$. In contrast, distributions of mother's time spent on child care differ marginally: $p_{MW} = 0.076$ and $p_{KS} = 0.099$. Regression results in Table 3.7 show that differences in the time that mothers spent on child care cannot explain different educational attainment of siblings when other explanatory variables are controlled for.

3.7.2 Robustness checks

Table 3.8: Robustness checks I

dependent variable: sibling difference in high secondary track attendance

specification sample dependent variable differenced variables	(1) general latest obs.	(2) West Germ. latest obs.	(3) general at age 14	(4) general latest obs.	(5) general latest obs.
mother's weekly hours	0.005	0.006	0.004	-	-
worked*	(0.105)	(0.113)	(0.268)	-	-
father's weekly hours	0.004	0.004	0.001	-	-
worked*	(0.529)	(0.629)	(0.892)	-	-
parents' joint weekly	-	-	-	0.004	-
hours worked*	-	-	-	(0.180)	-
parents' joint hours	-	-	-	-	-0.007
spent on child care*	-	-	-	-	(0.353)
male	0.009	0.006	-0.012	0.009	0.011
	(0.715)	(0.884)	(0.780)	(0.732)	(0.683)
born before July	-0.097	-0.112	-0.127	-0.094	-0.085
	(0.032)	(0.033)	(0.029)	(0.024)	(0.027)
firstborn child	0.151	0.151	0.165	0.148	0.168
	(0.003)	(0.057)	(0.007)	(0.002)	(0.001)
year of birth	-0.042	-0.043	-0.042	-0.043	-0.040
	(0.005)	(0.016)	(0.096)	(0.001)	(0.003)
household income**	0.019	0.074	0.379	0.030	0.174
	(0.961)	(0.867)	(0.585)	(0.935)	(0.715)
(household income)^2 **	0.026	0.006	-0.070	0.023	-0.022
	(0.837)	(0.986)	(0.748)	(0.849)	(0.883)
constant	0.317	0.330	0.374	0.317	0.327
	(0.000)	(0.050)	(0.000)	(0.000)	(0.000)
N	301	213	163	301	295
R^2	0.240	0.251	0.220	0.234	0.236

* average at ages 0-3 of child
** total monthly net equivalent income in 1000 Euros, average at ages 0-3 of child
comments: robust, clustered standard errors that allow observations to be correlated within a family; p-values are reported in brackets

Table 3.9: Robustness checks II

dependent variable: sibling difference in high secondary track attendance

key explanatory variables differenced variables	difference in weekly hours worked	difference in daily hours spent on child care
mother's time input at age 1	0.004	-0.001
	(0.082)	(0.860)
mother's time input at age 2	0.000	0.007
	(0.970)	(0.270)
mother's time input at age 3	0.000	-0.005
	(0.867)	(0.515)
father's time input at age 1	0.001	-0.002
	(0.754)	(0.891)
father's time input at age 2	0.003	-0.001
	(0.236)	(0.903)
father's time input at age 3	0.001	-0.008
	(0.824)	(0.708)
male	0.002	0.011
	(0.949)	(0.835)
born before July	-0.053	0.015
	(0.184)	(0.741)
firstborn child	0.164	0.150
	(0.010)	(0.018)
year of birth	-0.061	-0.058
	(0.000)	(0.000)
household income*	0.530	-0.147
	(0.234)	(0.807)
(household income)^2 *	-0.099	0.458
	(0.485)	(0.458)
constant	0.380	0.336
	(0.000)	(0.000)
N	219	108
R^2	0.326	0.416

* total monthly net equivalent income in 1000 Euros, average at ages 0-3 of child
comments: robust, clustered standard errors that allow observations to be correlated within a family; p-values are reported in brackets

Bibliography

[1] Ackert, L. F., Martinez-Vazquez, J., Rider, M., 2004. Tax policy design in the presence of social preferences: Some experimental evidence. Working Paper 04-25. International Studies Program, Georgia State University.

[2] Andreoni, J., Erard, B., Feinstein, J., 1998. Tax compliance. Journal of Economic Literature 36, 818-860.

[3] Andreoni, J., Miller, J., 2002. Giving according to GARP: an experimental test of the consistency of preferences for altruism. Econometrica 70 (2), 737-753.

[4] Andreoni, J., Vesterlund, L., 2001. Which is the fair sex? Gender differences in altruism. Quarterly Journal of Economics 116 (1), 293-312.

[5] Ashenfelter, O., Rouse, C., 1998. Income, schooling, and ability: evidence from a new sample of identical twins. Quarterly Journal of Economics 113 (1), 253-284.

[6] Bardsley, N., 2005. Dictator game giving: altruism or artifact? Working paper 1/2006. ESRC National Center for Research Methods, University of Southampton.

[7] Becker, G. S., 1968. Crime and punishment: an economic approach. Journal of Political Economy 76 (2), 169-217.

[8] Becker, G. S., Tomes, N., 1979. An equilibrium theory of the distribution of income and intergenerational mobility. Journal of Political Economy 87 (6), 1153-1189.

[9] Becker, G. S., Tomes, N., 1986. Human capital and the rise and fall of families. Journal of Labor Economics 4 (3) Part 2, 1-39.

[10] Beckman, S. R., Formby, J. P., Smith, W. J., Buhong, Z., 2002. Envy, malice and Pareto efficiency: an experimental examination. Social Choice and Welfare 19 (2), 349-367.

[11] Bénabou, R., Tirole, J., 2003. Intrinsic and extrinsic motivation. Review of Economic Studies 70 (244), 489-520.

[12] Bénabou, R., Tirole, J., 2006. Incentives and prosocial behavior. American Economic Review 96 (5), 1652-1678.

[13] Blank, R. M., 2002. Evaluating welfare reform in the United States. Journal of Economic Literature 40 (4), 1105-1166.

[14] Bohnet, I., Cooter, R. D., 2005. Expressive law, framing or equilibrium selection? Mimeo.

[15] Bowles, S., 2007. Social preferences and public policies: are good laws a substitute for good citizens? Working paper 496. University of Siena.

[16] Bolton, G. E., Katok, E., 1995. An experimental test for gender differences in beneficent behavior. Economic Letters 48 (3), 287-292.

[17] Bolton, G. E., Ockenfels, A., 2000. A theory of equity, reciprocity and competition. American Economic Review 90 (1), 166-193.

[18] Bond, D., Park, J.-C., 1991. An empirical test of Rawls' theory of justice: a second approach in Korea and the United States. Simulation and Gaming 22 (4), 443-462.

[19] Brickman, P., 1977. Preference for inequality. Sociometry 40 (4), 303-310.

[20] Buchel, F., Duncan, G. J., 1998. Do parents' activities promote children's school attainments? Evidence from the German Socioeconomic Panel. Journal of Marriage and the Family 60 (1), 95-108.

[21] Bundeskriminalamt (Ed.), 2005. Polizeiliche Kriminalstatistik Bundesrepublik Deutschland, Berichtsjahr 2005, Wiesbaden.

[22] Camerer, C., 2003. Behavioral game theory: experiments in strategic interactions. Princeton University Press, 43-117.

[23] Carlsson, F., Gupta, G., Johansson-Stenman, O., 2003. Choosing from behind a veil of ignorance in India. Applied Economic Letters 10 (13), 825-827.

[24] Carlsson, F., Daruvala, D., Johansson-Stenman, O., 2005. Are people inequality-averse, or just risk-averse? Economica 72 (278), 375-396.

[25] Chao, E. L., Rones, P. L., 2007. Women in the labor force: a databook. U.S. Department of Labor, U.S. Bureau of Labor Statistics.

[26] Croson, R., Gneezy, U., 2004. Gender differences in preferences. Mimeo.

[27] Curtis, R. C., 1979. Effects of knowledge of self-interest and social relationship upon the use of equity, utilitarian, and Rawlsian principles of allocation. European Journal of Social Psychology 9 (2), 165-175.

[28] Deci, E. L., 1971. Effects of externally mediated rewards on intrinsic motivation. Journal of Personality and Social Psychology 18 (1), 105-115.

[29] Deci, E. L., Koestner, R., Ryan, M. R., 1999. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. Psychological Bulletin 125 (6), 627-668.

[30] Dickinson, D. L., Tiefenthaler, J., 2002. What is fair? Experimental evidence. Southern Economic Journal 69 (2), 414-428.

[31] Dufwenberg, M., Muren, A., 2005. Generosity, anonymity, gender. Journal of Economic Behavior and Organization 61 (1), 42-49.

[32] Dustmann, C., 2004. Parental background, secondary school track choice, and wages. Oxford Economic Papers 56, 209-230.

[33] Dustmann, C., Schönberg, U., 2007. The effect of expansions in maternity leave coverage on children's long-term outcomes. Mimeo.

[34] Eckel, C. C., Grossman, P.J., 1998. Are women less selfish than men? Evidence from dictator experiments. Economic Journal 108 (448), 726-735.

[35] Eckel, C. C., Grossman, P. J., 2003. Men, women, and risk aversion - experimental evidence. In: Plott, S., Smith, V. (Eds.), Handbook of Experimental Results. Elsevier, forthcoming.

[36] Eide, E., 2000. Economics of criminal behavior. In: Bouckaert, B., De Geest, G. (Eds.), Encyclopedia of Law and Economics. Edward Elgar, Vol. V, 345-89.

[37] Ellingsen, T., Johannesson, M., 2008. Pride and prejudice: the human side of incentive theory. American Economic Review, forthcoming.

[38] Engelmann, D., Strobel, A., 2004. Inequality aversion, efficiency and maximin preferences in simple distribution experiments. American Economic Review 94 (4), 857-869.

[39] Ermisch, J., Francesconi, M., 2002. The effect of parents' employment on children's educational attainment. Working paper 2002-21. Institute for Economic and Social Research, University of Essex.

[40] Ermisch, J., Francesconi, M., 2005. Parental employment and children's welfare. In: Boeri, T., del Boca, D., Pissarides, C. (Eds.), Women at work: an economic perspective. Oxford University Press, 154-193.

[41] Falk, A., Fischbacher, U., 2002. Crime in the lab: detecting social interaction. European Economic Review 46 (4-5), 859-869.

[42] Fehr, E., Falk, A., 2002. Psychological foundations of incentives. European Economic Review 46 (4-5), 687-724.

[43] Fehr, E., Naef, M., Schmidt, K. M., 2006. The role of equality and efficiency in social preferences. American Economic Review 96 (5), 1912-1917.

[44] Fehr, E., Schmidt, K. M., 1999. A theory of fairness, competition, and cooperation. Quarterly Journal of Economics 114 (3), 817-868.

[45] Fehr, E., Schmidt, K. M., 2006. The economics of fairness, reciprocity and altruism: experimental evidence and new theories. In: Kolm, S.-C., Ythier, J. M. (Eds.), Handbook on the Economics of Giving, Altruism, and Reciprocity. Elsevier, Vol. I, 615-691.

[46] Fischbacher, U., 1999. Z-Tree: a toolbox for ready-made economic experiments. Working paper 21. University of Zurich.

[47] Fisman, R., Kariv, S., Markovits, D., 2007. Individual preferences for giving. American Economic Review 97 (5), forthcoming.

[48] Forsythe, R., Horowitz, J. L., Savin, N. E., Sefton, M., 1994. Fairness in simple bargaining experiments. Games and Economic Behavior 6 (3), 347-369.

[49] Francesconi, M., Jenkins, S. P., Siedler, T., 2006. Childhood family structure and schooling outcomes: evidence for Germany. DIW Discussion paper 610. Deutsches Institut für Wirtschaftsforschung.

[50] Frey, B. S., Jegen, R., 2001. Motivation crowding theory. Journal of Economic Surveys 15 (5), 589-611.

[51] Frey, B. S., Oberholzer-Gee, F., 1997. The costs of price incentives: an empirical analysis of motivation crowding-out. American Economic Review 87 (4), 746-755.

[52] Frohlich, N., Oppenheimer, J. A., Eavey, C. L., 1987. Choices of principles of distributive justice in experimental groups. American Journal of Political Science 31 (3), 606-636.

[53] Gächter, S., Kessler, E., Königstein, M., 2006. Performance incentives and the dynamics of voluntary cooperation: an experimental investigation. Mimeo.

[54] Galbiati, R., Vertova, P., 2005. Law and behaviours in social dilemmas: testing the effect of obligations on cooperation. Working paper 1/2005. University of Siena.

[55] Garoupa, N., 1997. The theory of optimal law enforcement. Journal of Economic Surveys 11 (3), 267-295.

[56] Glaeser, E. L., 1999. An overview of crime and punishment. Mimeo.

[57] Gneezy, U., 2003. The w effect of incentives. Mimeo.

[58] Gneezy, U., Rustichini, A., 2000a. A fine is a price. Journal of Legal Studies 29 (1), 1-17.

[59] Gneezy, U., Rustichini, A., 2000b. Pay enough or don't pay at all. Quarterly Journal of Economics 115 (3), 791-810.

[60] Graham, J. W., Beller, A. H., Hernandez, P. M., 1994. The effects of child support on educational attainment. In: Garfinkel, I., McLanahan, S. S., Robins, P. K. (Eds.), Child support and child well-being. The Urban Institute Press, 317-349.

[61] Haisken-deNew, J. P., Frick, J. R. (Eds.), 2003. DTC Desktop Companion to the German Socio-Economic Panel (SOEP). DIW, Berlin.

[62] Harsanyi, J. C., 1953. Cardinal utility in welfare economics and in the theory of risk-taking. Journal of Political Economy 61 (5), 434-435.

[63] Harsanyi, J. C., 1955. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. Journal of Political Economy 63 (4), 309-321.

[64] Haveman, R., Wolfe, B., 1995. The determinants of children's attainments: a review of methods and findings. Journal of Economic Literature 33 (4), 1829-1878.

[65] Hill, M. S., Duncan, G. J., 1987. Parental family income and the socioeconomic attainment of children. Social Science Research 16, 39-73.

[66] Holt, C. A., Laury, S. K., 2002. Risk aversion and incentive effects. American Economic Review 92 (5), 1644-1655.

[67] Houser, D., Xiao, E., McCabe, K., Smith, V., 2007. When punishment fails: research on sanctions, intentions and non-cooperation. Games and Economic Behavior, forthcoming.

[68] Irlenbusch, B., Sliwka, D., 2005. Incentives, decision frames, and motivation crowding out: an experimental investigation. IZA Discussion paper 1758. Institute for the Study of Labor, Bonn.

[69] Johannesson, M., Gerdtham, U.-G., 1995. A pilot test of using the veil of ignorance approach to estimate a social welfare function for income. Applied Economic Letters 2 (10), 400-402.

[70] Johansson-Stenman, O., Carlsson, F., Daruvala, D, 2002. Measuring future grandparents' preferences for equality and relative standing. Economic Journal 112 (479), 362-383.

[71] Kahneman, D., Knetsch, J. L., Thaler, R. H., 1986. Fairness as a constraint on profit seeking: entitlements in the market. American Economic Review 76 (4), 728-741.

[72] Kahneman, D., Tversky, A., 1986. Rational choice and the framing of decisions. Journal of Business 59 (4), 251-278.

[73] Kiernan, K., 1996. Lone motherhood, employment and outcomes for children. International Journal of Law, Policy and the Family 10, 233-249.

[74] Krupka, E., Weber, R., 2006. Why does dictator game sharing vary? Eliciting social norms in the laboratory. Mimeo.

[75] Leibowitz, A., 1974. Home investments in children. Journal of Political Economy 82 (2) Part 2, 111-131.

[76] Lepper, M. R., Greene, D., Nisbett, R. E., 1973. Undermining children's intrinsic interest with extrinsic reward: a test of the "overjustification" hypothesis. Journal of Personality and Social Psychology 28 (1), 129-137.

[77] Levitt, S., 1997. Using electoral cycles in police hiring to estimate the effect of police on crime. American Economic Review 87 (3), 270-290.

[78] List, J., 2007. On the interpretation of dictator game giving. Journal of Political Economy 115 (3), 482-439.

[79] Mitchell, G., Tetlock, P. E., Mellers, B. A., Ordónez, L.D., 1993. Judgments of social justice: compromises between equality and efficiency. Journal of Personality and Social Psychology 65 (4), 629-639.

[80] O'Brien, M., Jones, D., 1999. Children, parental employment and educational attainment: an English case study. Cambridge Journal of Economics 23, 599-621.

[81] Polinsky, A. M., Shavell, S., 2000a. The economic theory of public enforcement of law. Journal of Economic Literature 38 (1), 45-76.

[82] Polinsky, A. M., Shavell, S., 2000b. The fairness of sanctions: some implications for optimal enforcement theory. American Law and Economics Review 2 (2), 223-237.

[83] Puhani, P. A., Weber, A. M., 2007. Does the early bird catch the worm? Instrumental variable estimates of early educational effects of age of school entry in Germany. Empirical Economics 32, 359-386.

[84] Rawls, J., 1971. A theory of justice. Oxford University Press.

[85] Rosenzweig, M. R., Wolpin, K. I., 1995. Sisters, siblings, and mothers: the effect of teen-age childbearing on birth outcomes in a dynamic family context. Econometrica 63 (2), 303-326.

[86] Ruhm, C. J., 2004. Parental employment and child cognitive development. Journal of Human Resources 39 (1), 155-192.

[87] Schnepf, S. V., 2002. A sorting hat that never fails? The transition from primary to secondary school in Germany. Innocenti Working paper 92. UNICEF Innocenti Research Centre, Florence.

[88] Schulze, G. G., Frank, B., 2003. Deterrence versus intrinsic motivation: experimental evidence on the determinants of corruptibility. Economics of Governance 4, 143-160.

[89] Sliwka, D., 2007. Trust as a signal of a social norm and the hidden costs of incentive schemes. American Economic Review 97 (3), 999-1012.

[90] Solon, G., 1999. Intergenerational mobility in the labor market. In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics. Elsevier Science, Vol. 3, 1761-1800.

[91] Tamm, M., 2007. Does money buy higher schooling? Evidence from secondary school track choice in Germany. ECINEQ Working Paper 2007-58. Society for the Study of Economic Inequality, Palma de Mallorca.

[92] Todd, P. E., Wolpin, K. I., 2003. On the specification and estimation of the production function for cognitive achievement. Economic Journal 113 (485), 3-33.

[93] Torgler, B., 2002. Speaking to theorists and searching for facts: tax morale and tax compliance in experiments. Journal of Economic Surveys 16 (5), 657-683.

[94] Trautmann, S. T., 2007. Fehr-Schmidt process fairness and dynamic consistency. Mimeo.

[95] Tyran, J.-R., Feld, L. P., 2006. Achieving compliance when legal sanctions are non-deterrent. Scandinavian Journal of Economics 108 (1), 135-156.

[96] Waldfogel, J., Han, W.-J., Brooks-Gunn, J., 2002. The effects of early maternal employment on child cognitive development. Demography 39 (2), 369-392.

Curriculum Vitae

June 2004 - Februar 2008	Research and teaching assistant Chair of Prof. Dr. Klaus M. Schmidt Ph.D. program in economics Munich Graduate School of Economics Ludwig-Maximilians-Universität, Munich
March 2006 - August 2006	Visiting Ph.D. student in economics Stockholm School of Economics
October 2001- April 2004	Diplom in economics Universität Mannheim
August 2002 - May 2003	Visiting Ph.D. student in economics University of California, Berkeley
October 1999 - July 2001	Studies in economics, political and environmental science Universität Heidelberg
June 1999	Abitur Kurpfalz-Gymnasium Schriesheim
March 17, 1980	born in Mettmann, Germany