João Mira Canas da Eira

An Introduction to Other-Regarding Preferences With An Application to Contract Design

Dissertação de Mestrado em Economia, orientada pelo Prof. Doutor Paulino Maria Freitas Teixeira
e apresentada à Faculdade de Economia da Universidade de Coimbra

Janeiro 2018

· U C ·

UNIVERSIDADE DE COIMBRA

João Mira Canas da Eira

An Introduction to Other-Regarding Preferences With an Application to Contract Design

Coimbra, 2018

Para a Rita, que me atura, e para os meus pais, que me criaram, algo que a Rita pode ou não achar ter sido uma má decisão.

Resumo

Os modelos económicos de comportamento individual supõem, frequentemente, que na avaliação entre alternativas concorrentes os agentes apenas estão preocupados com a forma como cada alternativa os afeta pessoalmente. Esta simples e razoável suposição postula que os agentes cuidam apenas do interesse próprio (egoísmo racional), não se preocupando com o possível impacto das suas decisões sobre aqueles com quem interagem. O presente trabalho desafia esta suposição.

Ao longo das últimas décadas foi possível observar a acumulação de resultados experimentais provenientes de jogos como o ultimato e o *gift exchange*, onde o comportamento não é explicável com base em preferências puramente egoístas. Com efeito, os agentes geralmente tomam decisões que reduzem o seu bem-estar, desde que ao fazê-lo os restantes agentes possam beneficiar. Neste caso, em contraste com o puro interesse próprio, os agentes são ditos possuírem preferências sociais. Estes agentes estarão pois não são só preocupados com o que lhes acontece mas também com o que acontece aos outros agentes.

Uma larga parte dos resultados experimentais discutidos neste trabalho foram obtidos através do uso de experiências laboratoriais. A questão da validade externa destes resultados tem sido um ponto de disputa. As experiências laboratoriais são realizadas em ambientes altamente artificiais onde são colocadas fortes restrições no comportamento dos participantes. Embora isto lhes imbua com a sua fonte de força metodológica, é também uma fraqueza. Os resultados provenientes de experiências laboratoriais não generalizam necessariamente para o mundo real, e estes são frequentemente comparados com os resultados obtidos através de trabalho de campo devido à suposta maior validade externa destes últimos. A questão da validade externa das experiências laboratoriais é examinada e conclui-se que estas são uma ferramenta valida para a acumulação de evidência cientifica sobre o comportamento humano.

A aversão à desigualdade, é apresentada como um método de modelar preferências sociais. O modelo proposto é utilizado então para explicar o comportamento observado (em laboratório) no jogo do ultimato. Um exemplo sobre como utilizar preferências sociais, para estudar interações económicas no mundo real, é analisado na sua aplicação ao estudo da formulação de contratos sob risco moral.

Classificação JEL: D01; C70; B41.

Abstract

Economic models of individual behavior often make the assumption that in evaluating between competing alternatives agents are only concerned with how each alternative impacts their own payoffs. This simple, yet reasonable, assumption postulates that agents are self-regarding, that is, agents are not concerned with how their decisions affects other people. This study casts doubt over this assumption.

Over the last several decades there has been a steady accumulation of experimental evidence from games such as the ultimatum game and the gift exchange game where the observed behavior is not explained by assuming that agents have self-regarding preferences. Agents often make decisions that lower their payoff if by doing so other agents are better off. In contrast to self-regarding preferences, agents are said in this case to have other-regarding preferences. They are not only preoccupied with themselves but also with other people.

Most of the evidence discussed in this study was gathered through the use of laboratory experiments. The issue of the external validity of this evidence has long been a point of contention. Laboratory experiments are highly artificial environments that place strong constraints on individual behavior. While this imbues them with their source of methodological strength, it is also a weakness. Evidence gathered in the laboratory need not generalize to the real world, and laboratory experiments are often compared with field studies which purport to provide evidence that is more externally valid. We examine the question of the external validity of laboratory experiments and conclude they are a valid tool for the gathering of scientific evidence about human behavior.

Inequity aversion is presented as a method of modeling other-regarding preferences. The model is promptly used to explain the behavior documented in the ultimatum game. An example on how to use other-regarding preferences to study real world economic interactions is provided in the study of contract design under moral hazard.

JEL Classification: D01; C70; B41.

Table of Contents

1 Introduction

Economics is a *social* science and the economic behavior that is its subject of study is *human* behavior. As social scientists, economists are interested in studying agents, their actions, the reasons behind them, and the consequences that result from them. It is through the use of tractable models that economists perform their studies and, as a necessary step to develop these models, it is required that the goals and motivations that precede and drive human behavior be formalized.

We are regularly faced with situations where we have to choose between multiple possible courses of action. Before entering college we must decide between majoring in Physics or Economics. When lunch hour arrives, and we find ourselves at a mall, we have to choose one restaurant out from possible dozens. Economists deal with this basic fact of everyday life by introducing the concept of preferences, which are defined as rankings that express the subjective comparative evaluations of alternatives (Hausman, 2011).

An agent's behavior can be summarized as the maximization of an abstract utility function. While this utility function does not necessarily take into account the underlying psychological processes that underlie preference formation, it has become standard in Economics to take this function as being the result of an evaluation that takes into account as its sole parameter how each alternative impacts the agent's payoff. Under this behavioral assumption, the maximizing behavior corresponds to the idea that in the presence of competing alternatives people seek to maximize their own expected payoffs. Agents are then said to be self-regarding.

This assumption was put forth in 1881, when the political economist and philosopher Francis Edgeworth asserted that "the first principle of Economics is that every agent is actuated only by self-interest" (Edgeworth, 1881). More famously, we see it in Adam Smith's concept of the invisible hand, where the market is able to turn what are private vices into public virtues.[1]

This work intends to show that there is sufficient experimental evidence showing that the assumption that agents are self-regarding is insufficient. It will be shown that there is a wide range of behavior which the self-regarding assumption is not able to explain and that instead one needs to take into account that agents are concerned not only with themselves but also take

---

[1]Contrary to what one might infer from the Invisible Hand concept, Adam Smith never believed humans are only driven by self-interest (Smith, 1822).

into account the well-being of others.

We will therefore contrast *other-regarding preferences* with *self-regarding preferences*. An agent is said to be self-regarding if he is only preoccupied with how an action impacts himself, while an other-regarding individual is not only preoccupied with himself but also with other people.

Note that an individual with other-regarding preferences does not imply that he is not preoccupied with himself. For example, one can be honest because one does not wish to impose costs on others by deceiving them, but honesty can also be a self-regarding behavior if practiced in order to be the kind of person one wants to be. Thus, the distinction between the two preferences is not that other-regarding preferences are counter-preferencial, in the sense of behavior not following from the maximization of a utility function, but that agents are motivated by a concern about the effects of one's actions on others.

The present study is structured as follows. In Section 2 we will survey the evidence accumulated through the use of experimental games, such as the *ultimatum game* and the *gift exchange game*, that proves the existence of behaviors which the assumption of agents having self-regarding preferences is not able to explain. Section 3 provides a methodological defense against critics who argue against the use of laboratory evidence, such as that described in Section 2, to infer the determinants of human behavior. Following that, Section 4 introduces the Fehr-Schmidt model of inequity aversion, a model of other-regarding preferences that is able to predict the perplexing behavior documented in Section 2. Finally, in Section 5 we provide a motivating example for the use of other-regarding preferences by showing that their inclusion is able to explain the optimal choice between competing contracts under the existence of moral hazard.

2 Other-Regarding Preferences: Experimental Evidence

2.1 Ultimatum and Dictator games

2.1.1 The ultimatum game

The *ultimatum game* is a one-shot game between two players, a *proposer* and a *responder*. The *proposer* is given an integer amount of tokens, x, by the

experimenter and must offer a share of it to the responder. If the responder accepts, the proposer's offer is implemented and both part ways with their respective payout. If the responder rejects the offer both players part ways with nothing.

If both players have self-regarding preferences the proposer's optimal strategy will be to propose the lowest possible amount that he is allowed to offer. Accordingly, the responder should accept whatever amount the proposer is willing to part ways with, because otherwise he will be left with nothing rather than something. The subgame perfect Nash equilibrium for the ultimatum game is one where the payoffs are $(x - p, p)$, where p is the lowest possible amount that the proposer is allowed to offer. The experimental evidence, however, does not support this prediction.

Camerer (2011a) provides a detailed summary of the main results from a number of experiments using the ultimatum game. The main conclusions are as follows: The mean offer made by the proposer falls between 30% to 40% of the initial endowment. The median offer is 40% to 50%. There are rarely any unfair offers, that is, offers that fall in the 0% to 10% range. Offers that are too generous (i.e. more than half of the endowment) are also rarely observed. Low offers are often rejected, with offers below 20% being rejected about half of the time.

An increase in the stakes involved does not change the nature of the results. A possible objection might be that the stakes, or the monetary amount at stake in the interaction, are too low to elicit the required mental effort for players to play in the 'appropriate' manner. That is, if the stakes involved are low it is possible that players will not take the game seriously. However, when the stakes are increased players continue behaving in ways that do not conform to the self-regarding prediction.

For example, Cameron (1999) conducted experiments using the ultimatum game in Indonesia where the largest monetary amount at stake was equivalent to about three times the average monthly expenditure of the participants. The authors conclude in this case that "significant deviations from game-theoretic behavior persist even in high stakes games." The one change in player behavior that the authors were able to observe was that responders were willing to accept a lower percentage offer, while there was no behavioral change from proposers.

Andersen, Ertaç, Gneezy, Hoffman, and List (2011) employ the ultimatum game in a poor village in Northeast India to study the effect of an increase in stakes on responder behavior. They are motivated by the finding

that an increase in stakes does not elicit lower offers from proposers, which difficults the study of the effect an increase in stakes has in how responders deal with low, or unfair, offers. The authors increase the stakes by a factor of 1,000, — 20 to 20000 rupees (1.6 to 16000 hours of work) — and they alter the standard experimental instructions to elicit lower offers than usual from proposers. They find that responders play more closely to their predicted equilibrium response as stakes increase, usually as the amounts offered are equivalent to 30-40 days of wages or more. Rejection rates approach zero as the amount of money that responders must forgo with a rejection increases, meaning that stakes have their predicted effect. The authors point out that their finding confirms rather than rejects previous results given that one does not typically encounter situations where such high stakes are involved and the bulk of everyday market transactions are low-stakes affairs.

Slonim and Roth (1998) combine learning and increased stakes. Subjects from Slovak Republic play 10 rounds of the ultimatum game with stakes that between 60 and 1,500 Slovak crowns. Their results confirm previous findings, that behavior in the ultimatum game does not confirm the equilibrium predictions.

A possible objection might be that the observation of behavior not consistent with the self-regarding equilibrium prediction rises from the reliance of sterile laboratory experiments with college students, implying that the results do not generalize to a wider population. Early cross-cultural experiments with college students from Israel, the United States, Japan, and Yugoslavia, confirmed the standard finding in ultimatum experiments where the predicted equilibrium is never met, though the results did show substantial differences between countries regarding the distribution of offers made by the proposer (Roth, Prasnikar, Okuno-Fujiwara, & Zamir, 1991). These results provided some evidence that the deviation from the self-regarding prediction in the ultimatum game did generalize for populations all over the globe.

However, in 1996 a surprising finding broke the consensus when anthropologist Joe Henrich (Henrich, 2000) found that the Machiguenga, a slash-and-burn horticulturalist society living in the southeastern Peruvian Amazon, behaved in a way that was closer to the game-theoretic prediction. This "Machiguenga outlier" sparked the question of whether the behavior commonly observed in the ultimatum game was an artifact of the game being played by members of societies advanced in their economic development and propelled researchers to think about what economic and cultural circum-

4

stances made it so that the Machiguenga found the modal offer of 15% a fair offer.

The answer to these questions came when a group of 12 anthropologists, including Henrich, adapted the ultimatum game, the dictator game, and the public goods game so that these were not reliant on the administration through a computer and could thus be implemented in the field among non-literate subjects (Henrich et al., 2005). They proceeded to gather evidence from 15 small-scale societies exhibiting a wide variety of economic and cultural conditions.

In line with previous research, the predictions from the self-regarding model were not borne out in any of these societies, though there was wide variation in the results. The mean offers ranged from 26% to 57%, with the Machiguenga having the lowest mean offer and the Lamalera, a whale hunting people from near Indonesia, having the highest one. Indeed, the wide variation in how these societies approach the ultimatum game is quite interesting. The Hadza, a group of small-scale foragers from Tanzania, made low offers at the same time that they had a high rejection rate, while the Aché, from Paraguay, made consistently high offers with no rejections. The authors propose that this variation reflects their differing patterns of everyday life. Both groups share between members the meat that is obtained by hunters, though their levels of cooperation and expectations vary significantly. The Aché distribute their prey equally among all other households, and there is no consistent relationship between how much meat a hunter brings in and how much his family receives. Indeed successful hunters often leave their prey outside the camp to be discovered by others to avoid being considered boastful by their peers. By contrast, Hadza hunters sometimes wait until nightfall so they can sneak meat into their shelter, and when meat is shared between the group it is not done so without complaint and without some looking for opportunities not to share.

The authors reach the conclusion that increased sociality is dependent on the extent of the market integration in each society, that is, whether its people buy and sell wares and goods between one another and work for a wage. They find that increased cooperation in production is also associated with increased sociality, which might explain why the whale hunters of Lamalera feature such high levels of sociality, since it is necessary to sustain high levels of cooperation between multiple non-kin members to bring such an animal down. Taken these two aspects together, market integration and the payoffs to cooperation account for 66% of the variation in the outcomes in

5

the ultimatum game.

More amusingly, Carter and Irons (1991) find that economists play closer to the standard self-regarding prediction than non-economists. But there does not seem be a difference between freshman and senior economists. Economists, it seems, are just different from everyone else!
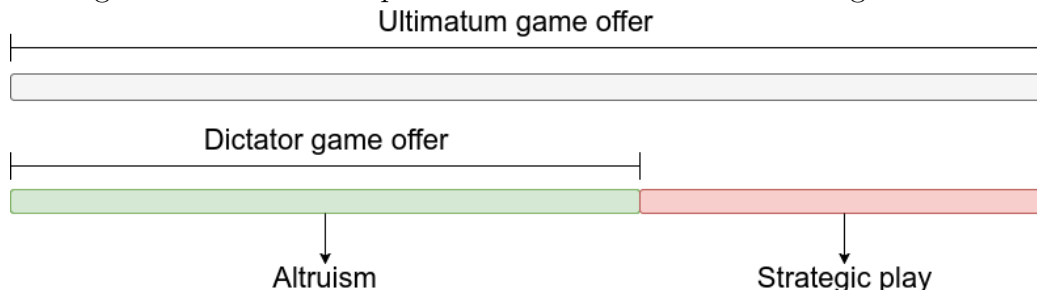
2.1.2   The dictator game

The *dictator game* is a variant of the ultimatum game where the responder is forced to accept the proposers offer regardless of the amount proposed. If the proposer has self-regarding preferences we would predict for him to propose the lowest denominator he is allowed to since there is nothing to be gained by offering a higher share of the endowment

Camerer (2011a) offers a summary of the results from multiple experiments that have employed the dictator game. The mean offer across these studies is roughly 20% of the initial endowment, and about 60% of the subjects in these studies offered a positive amount of the endowment.

That the proposer in the dictator game makes a mean offer that is higher than the minimum required, though lower than the mean offer in the ultimatum game, provides us with knowledge about the motives behind the offers made in the ultimatum game as well as the nature of those made in the dictator game itself. Given that the only meaningful difference between the dictator and ultimatum game is that in the first the ability of the responder to reject the offer made is removed, we can infer from the lower mean offer that strategic concern drives at least a portion of the offer made in the ultimatum game. That is, the proposer offers more than he would otherwise have offered due to a fear of his offer being rejected.

However, that the mean offer in the dictator game is not the minimum required tells us that this strategic concern does not entirely drive the offer in the ultimatum game. Given that the proposer is made worse off by offering more than minimum, and since the responder is a passive actor in this interaction, we can interpret the offer made in the dictator game as being driven by an aversion to inequality, or altruism.

Figure 1: The two components that drive the ultimatum game offer

List (2007) push back against the standard interpretation that positive offers in the dictator game reflect altruism and/or inequity aversion from the part of the proposer. For example, lower offers are seen when anonymity between proposer and responder is added, indicating that a concern about how one is seen by their peers is a driving force for the positive offers seen in the dictator game. List (2004) also find that, in the public goods game, the more anonymous decisions were amongst subjects the less the subjects opted to give in the one-shot version of the game.

Dana, Cain, and Dawes (2006) consider a variant of the dictator game where the proposer is given an initial endowment of $10 and, after having made his choice, is offered the option of exiting the game with $9. The exit option leaves the responder with nothing but ensures that he never knows that the game has been played. Even though proposers could get a higher payoff by engaging the receiver in a dictator game and not offering anything, 28% of the proposers opted for the exit option, perhaps because they didn't want to appear unfair to the receiver were they to enter the dictator game. In their second experiment the receiver never knows whether the money offered to them comes from the proposer or from the experimenters, thus allowing the authors to determine with more clarity whether appearing to be fair is indeed a concern for proposers. They find that 9 out of 24 proposers exited, which does imply a significant minority of proposers is concerned about not appearing self-regarding to the receivers.

In Dana, Weber, and Kuang (2007) a variant of the dictator game is played. Proposers are sorted into two different treatments, the baseline and the hidden payoff treatment. In the baseline treatment, the proposer can choose one of two actions, A and B, with respective payoffs $(6,1)$ and $(5,5)$

for the proposer and responder respectively. In the hidden payoff treatment the payoff for the responder is uncertain, so proposers must choose between actions A and B where the payoffs are shown to them as $(6, ?)$ and $(5, ?)$. All subjects are told that the payoffs from A and B are equally likely to be either $(i)\,(6, 1)$ and $(5, 5)$, or $(ii)\,(6, 5)$ and $(5, 1)$. The proposer can, at no cost to himself, choose to reveal the payoffs by clicking a button on the computer screen, and the responder is not made aware of that this choice has been made. The prediction is that if altruism is a better motivator for the proposer's actions the proposer will choose B in (i) and A in (ii).

By comparing the proportion of proposers that chose option B in the baseline treatment with the proportion in the hidden payoff treatment that chose to reveal the payoffs the authors are able to infer whether inequity aversion is an important motivation behind the offers.

They find 14 out of 19, or 74%, of proposers in the baseline treatment chose the more generous option B. However, in the hidden payoff treatment, 56% did not choose to click the button to reveal the payoffs, a difference in proportion that is statistically significant. These differences imply that the appearance of being fair is an important determinant in the offers made in the dictator game and that inequity aversion does not provide a full explanation. It is possible then that at least part of the positive offers in dictator games are made not because proposers are *altruists* but because they are *reluctant altruists*. They want appear to be altruists to everyone else but they would much rather keep the money to themselves.

2.2   Gift exchange and trust games

2.2.1   The gift exchange game

The gift exchange game was introduced by Fehr, Kirchsteiger, and Riedl (1993) in an attempt to empirically investigate whether the notion of fairness held by agents impeded the formation of a market clearing equilibrium in labor markets, a topic first broached by Akerlof (1982).

In the *gift exchange game*, two players are each assigned one of two roles: a firm or a worker. The firm offers a wage w to the worker, which the worker can then reject, in which case both earn nothing, or accept, in which case the worker must now expend an effort level, e, of his choice.

The standard prediction in such a setting can be discerned using a neo-classical model. Suppose the firm decides to offer the same wage to all its

workers, $\omega = \bar{\omega}$. Workers have a utility function, $u(\omega, e)$, where ω is the wage rate and e is the effort level they expend. The firm dictates that workers provide a minimum effort level in exchange for their wage, e_{min}. Workers, mindful of the firm's work rules, should choose their effort such that it maximizes:

$$u(\omega, e) \tag{1}$$

subject to the constraints

$$\omega = \bar{\omega} \tag{2}$$

and

$$e \geq e_{min} \tag{3}$$

This maximization problem yields the prediction that workers will choose the lowest effort level possible, e_{min}. The firm, aware of this, will set $\bar{\omega}$ as low as possible in an effort to maximize profits.

In his paper, George Akerlof is motivated to explore the effects of fairness in the formation of involuntary unemployment due to the curious results from a study of social relations among workers at a utility company in the eastern United States (Homans, 1954).

In this study, a group of women were found to be exceeding the minimum work requirements set by the firm by a considerable margin, a behavior that the neoclassical model above does not predict. Akerlof envisions this seemingly perplexing behavior as the result of the firm and the workers modelling their relationship as a "gift" exchange mediated by endogenous social norms. The workers offer a "gift" to the firm in the form of additional effort level, and in exchange the firm offers a "gift" in the form of a wage that the workers consider fair and that is in excess of what they could receive were they to leave their jobs. Thus, a labor market equilibrium is created where workers work harder because they are paid above opportunity cost. This wage level is higher than the market clearing one, ensuring that unemployment is present in equilibrium.

The gift exchange game permits us to study the level of *intrinsic reciprocity* in social relations such as the one described previously. This reciprocity falls in the category of other-regarding preferences.

Consider the following experiment from Fehr, Gächter, and Kirchsteiger (1997) Subjects were assigned into one of two roles: a principal or an agent.

Identities were kept anonymous, so no reputation building was possible. Principals make a job offer to the group of agents, meaning that principals stand in for employers and agents for workers. Agents are given the option to accept or reject the offer, and in an effort to spur competition there are more agents than principals. The job offer consists of an incomplete contract, (w_b, e_n), that specifies a *binding* wage level, w_b, and a *non-binding* effort level, e_n. The choice of the effort level is represented by the choice of a number in which the higher the number the higher the effort is, and the higher the costs borne by the agent are. Nothing in the experiment impedes agents from choosing an effort level that is lower than the proposed effort level in the contract as there is no punishment for doing so.

The expected behavior for both workers and firms are as noted earlier: agents will choose the lowest possible effort level and principals, knowing this, will offer the lowest possible wage level. However, if the principal believes there are sufficiently many reciprocal agents, he has an incentive to offer higher wages in an attempt to induce higher effort levels from the agents in reciprocity. Additionally, agents may induce reciprocity by the firms by offering a higher effort level than the one initially proposed.

Figure 2: Relation of desired and actual effort to the rent offered to workers
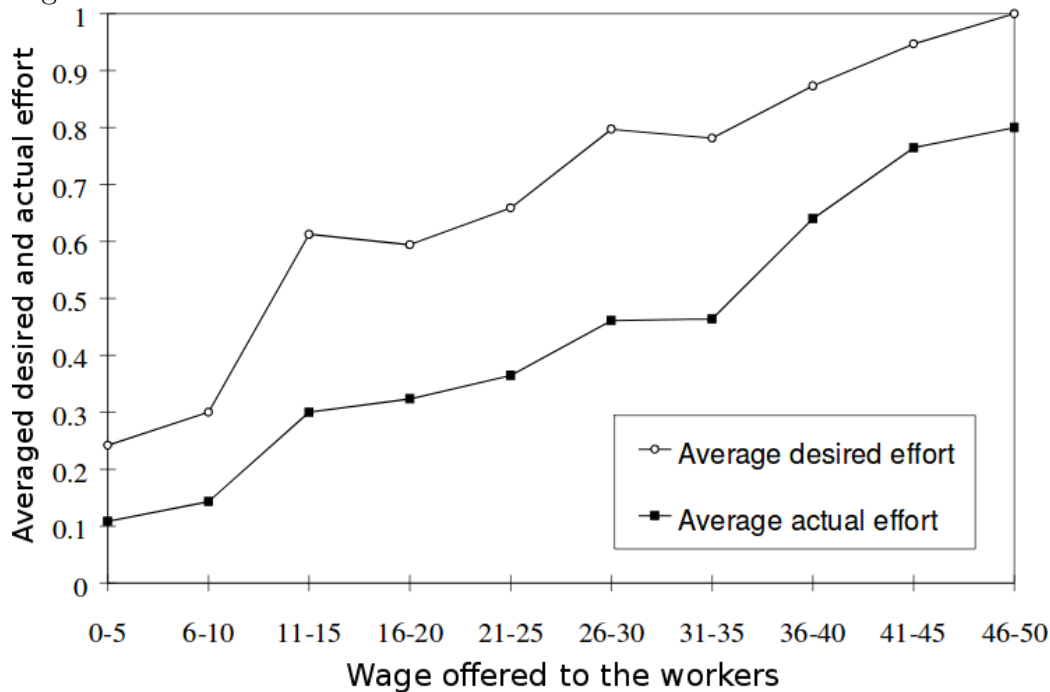
Source: Fehr and Falk (2002)

The experimental results are depicted in Figure 2. Two conclusions follow:

1. Higher desired effort levels are associated with more generous offers to the workers, which suggests employers try to elicit reciprocal responses from the workers

2. *On average*, the workers respond reciprocally to the employer's higher offers, though there is always a certain amount of shirking present.

The authors further add that *"there is also a substantial fraction of selfish workers who always choose the minimal effort or who rarely respond in a reciprocal manner."* The authors summarize the evidence from multiple studies to suggest that the fraction of self-regarding agents lies between 40% to 60%.

While we will take up the issue of the external validity of laboratory experiments in a later section, it is worthwhile to point out some of the pushback

against the main conclusions of the gift exchange game that have arisen from the results gathered from the use of field studies to study reciprocity.

Gneezy and List (2006) hired students to a data-entry job where they would enter books into a library information system. Each student performed the task alone, and were offered $12 for the job. In the experimental treatment, after the training phase, a portion of the students were informed they would receive $20 per hour, with no explanation for the increase in pay. Students in the control condition were payed the previously agreed upon $12.

The results seemed to cast doubt over the idea that offering a wage premium is an effective measure to elicit higher worker performance. In the first 90 minutes, those workers in the treatment condition produced around 25% more than their control peers. Although this percentage difference in effort is noteworthy, the increase in effort vanished as the experiment continued and effort levels for both treatment and control conditions where found to not be significantly different. The authors interpret their results as showing that while higher wages are reciprocated by greater effort on the part of the workers, this higher effort is not persistent and thus we need be careful to extrapolate from the single round interactions featured in the gift exchange game to how these relationships actually develop in the real world.

A problem with Gneezy and List (2006) is that of a small sample size, which limits their ability to detect statistical significance if the effect of a wage premium on effort is modest or small, a point mentioned by Fehr, Goette, and Zehnder (2009). Indeed, Cohn, Fehr, and Goette (2008) use a larger sample size, which allows them to have enough power to detect a statistically significant increase in effort from the increased wage, not replicating Gneezy and List (2006). Fehr et al. (2009) surveys the literature and concludes the positive relationship between wage and effort to be robust and well replicated.

2.2.2 The trust game

The *trust game* is played between an *investor* and a *responder*. Each player is endowed with a fixed amount of tokens, x. The investor must decide an amount $i \leq x$ to send to, or invest with, the responder. Before the amount chosen is delivered to the responder, the experimenter multiplies it by a multiplier m, meant to capture market return, and passes it on to the responder. The responder must then return an amount $r \leq mi$ back to the investor.

If both subjects have self-regarding preferences then the responder will

never send any money back to the investor. The investor, correctly anticipating the responder's behavior, will decide not to invest any amount i.

Figure 3: The trust game
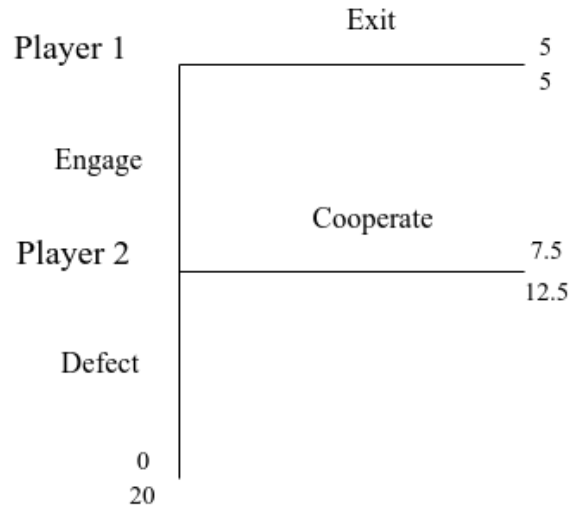
We can use a concrete example to prove this prediction. Let us assume a game with two players, Player 1 and Player 2. Both are endowed with $5. Player 1 can decide between keeping his endowment, in which case the game is ended and both players walk off with a payoff of $5, or he may pass the entire endowment to Player 2. If the latter, the endowment is tripled by the experimenter and it is then up to Player 2 to decide whether to keep the additional $15 for himself, or return, for example, $7.5 to Player 1. The payoffs are, respectively, ($0, $20), and ($7.5, $12.5)

The subgame perfect Nash Equilibrium of the trust game for the self-regarding preferences model can be determined using backward induction. In the second stage of the game, Player 2 maximizes his payoff by defecting and walking off with the full amount. Predicting this, Player 1 will not send his endowment to Player 2. Thus, the predicted payoff will be ($5, $5), that is, both players walk out with their initial endowment, having not cooperated. Traditionally, trust game experimenters allow for both players to choose how much they intend to send to the other player, but this does not change what the subgame perfect equilibrium is.

13

The share of the endowment the investor decides to invest is said to capture trust and the share sent by the responder, thrustworthiness. Both are forms of other-regarding preferences.

If both players have self-regarding preferences, then the subgame perfect equilibrium will be met. However, if both players have other-regarding preferences we should see a positive amount sent by the investor, $i > 0$, and a positive amount returned by the responder, $r > 0$. Figure 4 shows the distribution of offers made by both the investor and responder in a meta-analysis of 161 studies involving approximately 24,000 participants (Johnson & Mislin, 2011).

Figure 4: Distribution of percentages sent by investors (left) and responders (right).
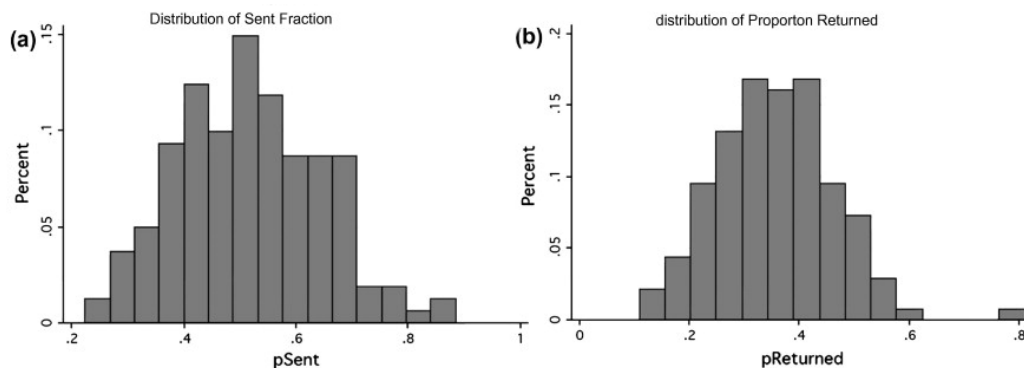
Source: Johnson and Mislin (2011)

The meta-analysis found that the mean offer made by the investor across all studies is .502, around half of the initial endowment, while the mean amount returned by the responder is .372. Due to aggregation of multiple experiments from multiple parts of the world the authors were able to study the differences in how people from around the world play the trust game. They find Africa sends and receives the lowest amount of all continents, with North America and Europe featuring the highest amounts both sent and received. They find further that older people send larger amounts, students send significantly lower amounts than non-students, and that amounts sent are larger if the subject believes he is playing with another human player.

3 On the Validity of Using Experiments in Economics

"One possible way of figuring out economic laws ... is by controlled experiments... Economists [unfortunately]... cannot perform the controlled experiments of chemists or biologists because they cannot easily control other important factors. Like astronomers or meteorologists, they generally must be content largely to observe"

Samuelson and Nordhaus (1985)

Laboratory experiments are a widely used tool in the physical and life sciences. In contrast, the social sciences have traditionally been considered nonexperimental, that is, the data upon which social scientists base their theories are collected not through experimentation, but observation. This is due to the obvious constraints of this class of scientific disciplines (e.g. historians are not able to recreate in a lab the Napoleonic Wars). This is not to say the social sciences have not or do not use laboratory experiments when it is feasible to do so. Psychology, for instance, has used laboratory experiments, or the experimental method more broadly, successfully for more than two centuries now (e.g. Ebbinghaus (1885)). In economics, however, the use of laboratory experiments is a recent development.

Although there was some proto-experimental work done in the 1930s on the topic of consumer demand theory (Moscati, 2007), it is generally agreed that, as an institutional and intellectual programme, experimental economics took form in the late 1940s following the publication of John von Neumann and Oskar Morgenstern seminal *Theory of Games and Economic Behavior* in 1944 (Guala, 2008). Since then, the growth of published papers using laboratory experiments has been remarkable.

In three of the most prestigious economics journals — *American Economic Review, Econometrica*, and *Quarterly Journal of Economics* — the fraction of experimental papers published in proportion to all published papers was between 0.84% and 1.58% in the 1980s, jumping to 3.8% and 4.15% between 2000 and 2008 (Falk & Heckman, 2009). The first specialty journal, aptly named *Experimental Economics*, was founded in 1998. Moreover, 6% of the Nobel economics prizes awarded since 1969 have been to economists who can be described as working in experimental economics, in-

cluding heavyweights such as Elinor Ostrom, Daniel Kahneman, and, more recently, Richard Thaler. The rise in prominence of experimental economics has been such that, despite the above quote from Samuelson and Nordhaus, in the 1992 revision of their famous textbook they saw the need to further add that experimental economics is "an exciting new development"(Samuelson & Nordhaus, 1992, p. 5).
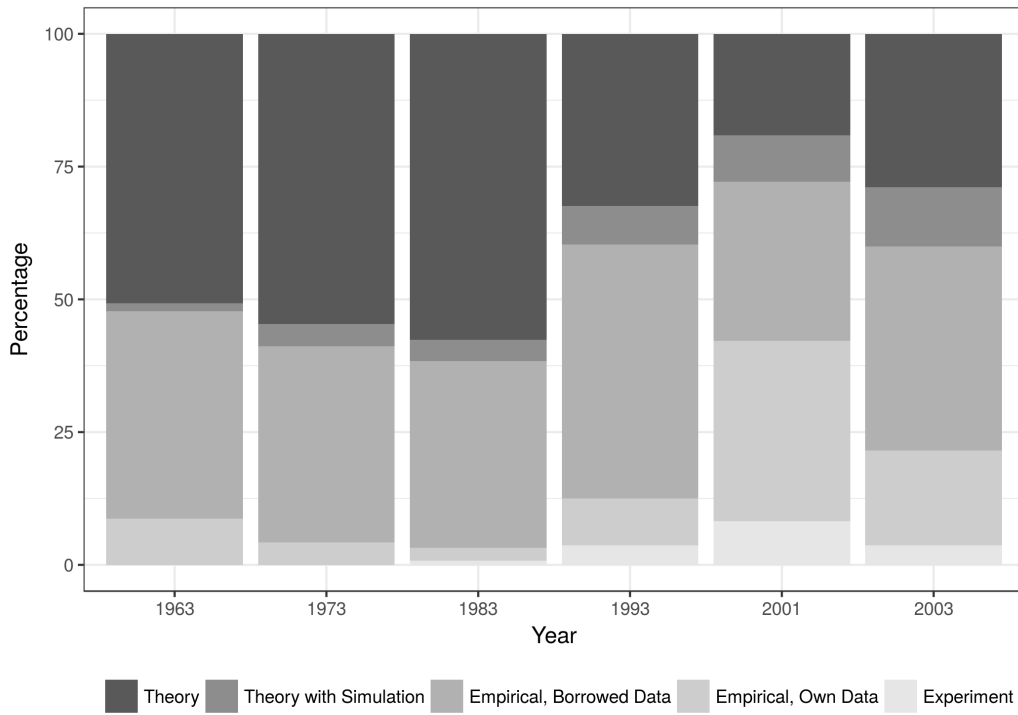
Figure 5: Methodology of articles in top economics journals, as percent of total. Source: (Mearsheimer & Walt, 2013)

There is however an elephant in the room. Despite the evident growth in the use of laboratory experiments in Economics, there have been challenges regarding whether the results coming out from these experiments allow researchers to say anything about economic behavior outside the lab. The issue of external validity, the ability of experiments to provide findings that are likely to allow for reliable inferences outside the laboratory, is a pernicious problem for the social sciences that does not exist to the same extent in the physical sciences.

By way of illumination, a Physics student performing a careful experiment to determine the value for Earth's gravity need not concern himself as to whether his results will generalize to outside the lab. The same is not necessarily true in the social sciences, in general, and for most of the experiments we have surveyed above, in particular. While it does seem that subjects offer around 20% of their endowment in the dictator game, we do not regularly see strangers in the streets spontaneously offering a fifth of the contents in their wallets to strangers passing by.

If we are to use the extensive evidence surveyed previously to argue for the existence of other-regarding preferences we must first establish that it offers us reliable evidence that extends beyond the artificial conditions of the laboratory. Indeed, it is this artificiality that, while imbuing laboratory experiments with their unique methodological strength, also gives it the weakness that has been an influential source of skepticism about their use as a tool in the economist's toolbag.

Laboratory experiments are often contrasted with field experiments. In the debate between the two, field experiments are often touted as possessing more 'realistic' conditions, even if they are perhaps less tightly controlled.

List (2006) provides an interesting example of this idea. In his paper, a gift exchange game is played where buyers make price offers to sellers and, in return, sellers select the quality of the good they will exchange with the buyer. The experiment was run in a standard laboratory setting and used experienced sports-card traders as subjects. The results mirrored the typical findings for this type of game: in the presence of higher offered prices, sellers tended to offer higher quality goods in return, even though they were not obligated to do so by the rules of the experiment. Thus this laboratory experiment points toward the existence of other-regarding preferences in the gift exchange game.

The experiment was then carried out by making a single change in which the goods exchanged where actual baseball cards whose market value was influenced by differences in their physical condition that experienced card sellers are more likely to detect than unexperienced sellers. In this experiment, other-regarding preferences were also observed. Higher quality cards were offered to buyers who offered higher prices.

The two experiments are therefore concordant in their conclusion of the existence of other-regarding preferences. However, List did not stop there. He wanted to know whether his results would also be observed in the card sellers natural environment, a sports-card show. Dealers in this field study were

unaware that their behavior was being observed and studied. Confederates were instructed to approach sellers and offer different prices in return for sports-cards of varying quality, mirroring the methodology of the previous experiments.

In this field study, where the dealers did not believe that the consumers could grade the cards appropriately or there was little possibility of future interaction, little statistical relationship between price and quality was observed. Only when there was concern for one's reputational standing, when sellers expected future interactions to happen or buyers could verify the quality of the cards by using a third-party, was high offered price met with high quality offered. Thus, the other-regarding preferences routinely observed in the lab were attenuated or not observed in the field condition.[2]

The dichotomy of results between the laboratory and field conditions should certainly make us pause before claiming with certainty that other-regarding preferences exist. It is tempting to think that field experiments reflect a more realistic condition and should thus be held in higher regard when making inferences about the inexistence of other-regarding preferences. Thus, that card dealers do not seem to match quality with price in a sports-card show seems to suggest that we should not change our priors with regards to the existence of other-regarding preferences. It is however worth thinking more deeply about this notion that field studies offer us a more realistic picture than laboratory experiments.

Camerer (2011b) makes the helpful distinction between the *policy view* and the *scientific view*. In the policy view, the generalizability of experimental findings to the outside world, that is, their *external validity*, is of paramount importance. This is because in choosing what policy to apply, evidence that has been collected in the same domain as the policy has obvious advantages in the inferences one can draw from it with respect to the effects of policy in question. Field experiments should be given more weight in this view than laboratory experiments do. For the scientific view, however, both laboratory and field studies constitute ways to enhance our understanding of human behavior and should therefore hold equal weight in the inference process. Provided the evidence was properly gathered and is valid and contextually relevant, there is no hierarchical relationship between

---

[2]Camerer (2011b) provides a critical reply to List (2006) and reanalyzes the data with different statistical techniques. He notes that "...the conclusion that the lab and field show different reciprocity is suggestive but is just not robustly significant."

the two methodologies. Both constitute tools to be used in the accumulation of knowledge. As Camerer puts it: "in this view, since the goal is to understand general principles, whether the 'lab generalizes to the field' ... is distracting, difficult to know ..., and is no more useful than asking whether 'the field generalizes to the lab'."

To illustrate this point, consider the following formalization from Falk and Heckman (2009). Suppose a variable of interest Y can be fully explained by the variables $X_1, ..., X_N$ and that the functional relationship between them is given by $Y = f(X_1, ..., X_N)$, known as the all-causes model. Suppose we are interested in examining the causal effect of X_1 on Y, which requires us to hold all other variables $\hat{X} = (X_2, X_3, .., X_N)$ fixed.

In a laboratory experiment the researcher estimates a model of the form $Y = f(X_1, X^L)$ where $X^L \neq \hat{X}$. Following the same logic, in a field experiment one estimates a model of the form $Y = f(X_1, X^{F_1})$ where $X^{F_1} \neq \hat{X}$ and typically $X^{F_1} \neq X^L$. The claim is usually made that $f(X_1, X^L)$ does not satisfy external validity but $f(X_1, X^{F_1})$ does.

We can write the field study in List (2006) as $Y = f(X_1, X^{F_1})$ where Y is the quality of the cards, X_1 is the price offered, and X^{F_1} are the remaining variables. Suppose we were to repeat that same experiment using a different subject pool, possibly stamp collectors. This gives rise to yet another estimated relation $Y = f(X_1, X^{F_2})$ where X^{F_2} reflects the set of variables and characteristics in this new experiment, including the new subject pool. How are we to adjudicate between $f(X_1, X^L)$ and $f(X_1, X^{F_1})$ if we want to predict the causal relationship between X_1 and Y in the new relation $Y = f(X_1, X^{F_2})$?

One might be tempted to reply that the field experiment should hold more weight than the laboratory experiment in this particular inference since the new situation also involves a field experiment. There is not, however, an explicit reason for why this should be so. While both situations are field experiments, there is no a priori reason to expect the behavior from sports-card dealers to generalize toward that of stamp collectors. As Camerer (2011b) puts it (using different notation), "if the litmus test of 'external validity' is accurate extrapolation to X^{F_2}, is the lab X^L necessarily less externally valid than the field setting X^{F_1}? How should this even be judged?"

It is best then, under the scientific view, to treat laboratory and field data as complementary. Both have their strengths, and the usefulness of one versus the other is ultimately a matter of the underlying research question.

Laboratory experiments, due to their tight control, are more prone to be replicable whereas in field experiments replicability can be challenging and is more often than not impossible. The smaller cost of laboratory experiments, as well as the easiness of archiving and reproducing instructions, software, recruiting protocols, databases, and statistical tools, also make laboratory experiments easier to replicate.

The two methods differ in what variables $\hat{X}$ they are able to provide a larger variation for. Laboratory experiments can explore the parameter space for values that can be hard or rare to find in the field. For example, in Andersen et al. (2011) the authors raise the stakes in the ultimatum game by a factor of 1,000 such that the highest stakes in play equal 16,000 hours of work, an amount that would be nigh impossible to find in the field. Field experiments do a better job at collecting evidence for different subject pools with different demographics and social characteristics, as we saw in the multidisciplinary work of Henrich et al. (2005) who employed the ultimatum, dictator, and public goods game in 15 small-scale societies.

Having argued for the usefulness of laboratory experiments, there still remain some concerns that need be addressed before we can be safe in using the experimental evidence to argue for the existence of other-regarding preferences. While we have seen that there is no a priori reason for laboratory experiments to not be used in the making of inferences about economic behavior in lieu of field experiments, this is only so if laboratory experiments are a valid tool for the study of this particular subject. That is, while by themselves laboratory experiments are a valid tool in the economist's toolbox, there might be flaws in the experimental process that invalid the use of the results gathered to infer the existence of other-regarding preferences. A powerful critique of the experimental method is offered in Levitt and List (2007) where the authors raise legitimate concerns that put into question the external validity of the observed results.

By the order in which these objections will be tackled, they are:

1. There is unprecedented experimental scrutiny in laboratory experiments. This may give rise to *experimenter demand effects* where subjects, perceiving that they are being observed by the experimenter, may behave in ways that they believe the experimenters desire, or may also behave in ways that end up not revealing their true preferences, e.g., being observed may lead to more prosocial behavior than the subject actually desires.

2. Human behavior is context-dependent and it is not clear that experiments can either capture or control this.

3. There is the possibility of self-selection bias where experiments might be being run with an homogeneous sample of students who might be more prosocial, more educated, and have a higher need for approval than the average human population.

Levitt and List additionally question the common use of small stakes in laboratory experiments arguing that it might not capture the richness of human behavior. We choose not to deal with this objection because the previous section already dealt with the effects of varying stake size in the experiments. Indeed, the issue of stake size has long been a topic of interest in behavioral economics (Camerer, Hogarth, Budescu, & Eckel, 1999) and is no longer a novel nor potent objection.[3]

Experimenter demand and audience effects

Subjects in laboratory experiments know that their behavior is being recorded and will be under intense scrutiny. This intense obtrusiveness might lead to subjects matching their behavior with what they perceive to be the experimenter's desired behavior (*experimenter demand effect*), or may lead subjects to behave in more prosocial ways because they believe that self-regarding behavior may be frowned upon by the experimenter (*audience effect*).

Let us assume for a moment that subjects hold an accurate view of what the experimenters expect and favor a particular outcome, which is not obvious and is something that experimenters are aware of when designing experimental procedures and thus work to circumvent. For there to be a demand effect it is necessary that the subject be willing to sacrifice his earnings by behaving in the way that the experimenter desires. Even if the subject is willing to do so, that willingness is but a component of his overall preference bundle, meaning that it should be possible to devise a situation that stresses his initial desire and makes him more reluctant to sacrifice his earn-

---

[3]As Colin Camerer writes in his 2003 textbook, five years before Levitt and List's paper: "If I had a dollar for every time an economist claimed that raising the stakes would drive ultimatum behavior towards self-interest, I'd have a private jet on standby all day (Camerer, 2011a, p. 60)."

ings. That increasing the stakes involved in the experiments typically has little effect suggests that demand effects are not a strong concern.

This still leaves us with the possibility that the intense experimental scrutiny leads to subjects behaving more prosocially. Barmettler, Fehr, and Zehnder (2012) employ a novel experimental procedure that allows the manipulation of experimenter-subject anonymity and employ it in three experimental games: the ultimatum, the dictator, and the trust game. In none of these, for any player in any role, is there a statistical significant difference in the choices made between the treatment condition where anonymity between subject and experimenter is maintained and the one where it isn't.

Experimenter demand and audience effects do not seem to be strong objections against the experimental evidence. Indeed, if experimental scrutiny made such a noticeable effect on the emergence of other-regarding preferences that would itself be evidence for the existence of other-regarding preferences for we are regularly being directly observed in many of our real-world interactions. More distopically, if scrutiny was a powerful enough force to flip a person's switch towards behaving more prosocially, it is likely we would be seeing intense institutional efforts to promote that scrutiny.

Context dependence

Human behavior is embedded in a rich, complex, and tangled web of social norms, frames, and the lessons learned from past experience. Levitt and List argue that different experimental procedures, such as writing the instructions the subjects read in different ways, may lead to differing results. Defection rates in the prisoner dilemma game vary depending on whether subjects are playing a "Community" or "Wall Street" game (Ward, 1997), framing the allocation of funds in a public goods game as a "contribution" or "allocation" matters, as does whether the game is framed as a positive externality or a negative one (Andreoni, 1995).

However, even though Levitt and List make a persuasive case for the importance of the context, they also argue that this context is "is not completely controlled by the experimenter." This collides with their previous argument because if it is possible to elicit different results by varying some of the experiment's parameters then it follows that it is also possible to control and account for that context. Indeed, that variation is highly desirable in the scientific view since it allows for the accumulation of knowledge about the boundary conditions of human behavior. Thus the very idea that experimen-

tal context might influence the results is an argument for more experiments to be run.

Moreover, to the extent that there is a subset of context that is not liable of being captured by any methodological variation, then it is also unlikely that uncaptured context is possible to control for in a field experiment. The problem of uncontrolled variables that cannot be measured and controlled for is not a problem of laboratory experiments per se, for field studies also run into it, but is instead a problem that every empiricist must wrestle and contend with.

Self-selection bias

The empirical evidence we have surveyed is largely based on laboratory experiments using self-selected students. This homogeneous sample might constitute a problem if students are found to behave in systematically different ways than the rest of the population. If the students who self-select into experiments behave more prosocially then the results provide a biased estimate of the extent to which there are other-regarding preferences in the population.

Falk, Meier, and Zehnder (2013) provide an interesting study of this question where they are able to distinguish whether students who self-select into laboratory experiments are any different from those students who don't, as well as whether these have different social preferences from the rest of the population.

They analyze the decisions of a sample of 16,666 undergraduates at the University of Zurich for which they know who participated in experiments and how many times. To measure the extent of their prosocial inclinations they use data collected from a naturally occurring repeated decision where each student must decide whether or not he or she wants to contribute a predeterminate amount to two social funds which provide charitable services. The authors conclude that participating students do not have different social preferences than their nonparticipating colleagues. If there is a bias then it is because students are different from everyone else.

To see if such a difference exists the authors employ two identical trust games using distinct subject pools so that the only difference in prosocial behavior comes from differences between the two subject pools employed and not changes in experimental design. One group was recruited from the student pool at the University of Zurich and the other from a representative

sample of the population of the city of Zurich. In total 1,296 participants were recruited (295 students and 1,001 from the general population)

They find that the non-student subject pool behaves more prosocially than the student subject pool which implies that by regularly employing students in their experiments researchers might be downwardly biasing their inferences about the existence of other-regarding preferences, i.e., to the extent that this is an issue is more one of magnitude than direction. The finding that students exhibit less prosocial behavior than the rest of the population is consistent with a range of other evidence, e.g., CEOs tend to be more prosocial than students (Fehr & List, 2004).

Indeed, concerns about the homogeneity of the subject pools used in laboratory experiments being problematic has been dealt with powerfully with the experiments performed by Henrich et al. (2005). Even though they report wide variation in the extent to which people in different societies have other-regarding preferences, the fact that none of the societies studied confirmed the self-regarding predictions implies that other-regarding preferences may well be a general feature of human nature.

4   Moddeling Other-Regarding Preferences

The evidence surveyed previously establishes that agents often have preoccupations not only about what happens to themselves but also with what happens to other people. Economic models often do not include these social preferences in their structure, possibly limiting the set of behaviors they are able to explore. At worst, by not taking into account the existence of other-regarding preferences these models may reach incorrect conclusions about the economic behavior of agents.

There is thus the need for a theoretical model that takes into account the empirical findings on other-regarding preferences. The Fehr-Schmidt model of inequity aversion developed in Fehr and Schmidt (1999) is a proposal for such a model. A description of this model and an application to the ultimatum game follows.

4.1   The Fehr-Schmidt Model of Inequity Aversion

Consider n individuals, each with a respective monetary payoff $y_1, y_2, ..., y_n$. The payoffs of all individuals but the individual i is denoted by the vector $\vec{y}_{-i}$.

For any i, the Fehr-Schmidt utility function, henceforth FS utility function, is defined as:

$$U_i \left(y_i, \vec{y}_{-i}; \alpha_i, \beta_i\right) = y_i - \frac{\alpha_i}{n-1} \sum_{j \neq i} \max\left\{y_j - y_i, 0\right\}$$
$$- \frac{\beta_i}{n-1} \sum_{j \neq i} \max\left\{y_i - y_j, 0\right\} \tag{4}$$

where $\alpha_i \geq 0$ and $0 \leq \beta_i < 1$. It is easy to see that the FS utility function describes an individual with self-regarding preferences if $\alpha_i = \beta_i = 0$, in which case $U_i \left(y_i, \vec{y}_{-i}; \alpha_i, \beta_i\right) \to U_i \left(y_i\right)$.

The FS utility function models the individual as comparing his situation with the situation of the individuals around him. That is, the individual exhibits self-centered inequity aversion, where some people are better off than him and he his better off than other people. This is captured by the second and third terms in the function, respectively.

The second term in the FS utility function measures the utility loss from disadvantageous inequality, more coloquially called envy, while the third term measures the loss from advantageous inequality, or altruism. It is assumed that $\alpha_i \geq \beta_i$. This means that, for the same magnitude, an individual loses more utility from another individual being better of than him than in the contrary situation. Envy is a more psychologically salient condition than altruism.

While there is no upper bound on α_i, it is assumed that $\beta_i < 1$. The reasoning for this upper bound is that when an individual with $\beta_i \geq 1$ reduces his advantage over other individuals he is increasing his utility by more than he is reducing his advantage. This seems implausible as it would require extremely high levels of altruism from the individual. Eckel and Gintis (2010) report the magnitudes of α and β from various studies. The evidence indeed suggests that for most individuals $\beta < 1$ and $\beta < \alpha$.

We will now see how the Fehr-Schmidt model can be used to understand the behavior seen in the ultimatum game that was previously unexplainable using models with just self-regarding preferences.

4.1.1 The ultimatum Game

A proposer and a responder play the ultimatum game. They are, respectively, Player 1 and Player 2. Of the full endowment researchers award to Player 1,

the share of it that is proposed is denoted by s.

Lemma 4.1. *The responder accepts all offers $s \geq 0.5$. There is a critical share, $s_c < 0.5$ such that the responder rejects all offers below it and accepts all offers $s \geq s_c$.*

Proof. If $s > 0.5$, taking into account the FS utility function for the case with only 2 players, we have the following for the responder:

$$U_2 = s - \beta_2 \left[s - (1 - s) \right] \tag{5}$$

which is positive because $\beta \in [0, 1)$, hence the responder will accept.

Now suppose $s < 0.5$. In this case we have

$$U_2 = s - \alpha_2 \left[(1 - s) - s \right] = s(1 + 2\alpha_2) - \alpha_2 \tag{6}$$

For this to be positive we need s such that

$$s \geq \frac{\alpha_2}{1 + 2\alpha_2}$$

Taking $\alpha \to \infty$ reveals that the critical threshold, s_c, is 0.5. $\qquad\square$

Proposition 1. *The equilibrium share offered by the proposer is given by:*

$$s^* = \begin{cases} s_c & \text{if } \beta_1 < 0.5 \\ 0.5 & \text{if } \beta_1 > 0.5 \\ s \in [s_c, 0.5] & \text{if } \beta_1 = 0.5 \end{cases} \tag{7}$$

From the previous lemma we know that the responder will accept any share $s_c \leq s \leq 0.5$. Let us consider such a share. From the FS utility function we have, for the proposer, $U_1 = (1 - s) - \beta_1 \left[(1 - s) - s \right]$. Taking the first derivative with respect to s leaves us with $2\beta_1 - 1$. Thus, if $\beta_1 < 0.5$, we have $\frac{\partial U_1}{\partial s} < 0$ so the proposer should offer the minimum possible that the responder will accept, i.e., s_c.

If $\beta_1 = 0.5$ we have $\frac{\partial U_1}{\partial s} = 0$ so any feasible share between s_c and 0.5 may be offered and will be accepted.

For values of B_1 higher than 0.5, $\frac{\partial U_1}{\partial s} > 0$, so we have $s^* = 0.5$.

Given what we know about the results usually observed from playing the ultimatum game, we can see that the Fehr-Schmidt model matches the experimental results reasonably well. Fehr and Schmidt note that their model

"yields too extreme predictions in ... the dictator game", where proposers only offer high offers ($s = 0.5$) or very low offers ($s = 0$), a prediction rejected by the data.

They note that their model can easily be modified so that the assumption that inequity aversion is linear can be dropped and substituted by a utility function that is concave in the amount of advantageous inequality. With this new assumption the results of the dictator game can also be accommodated.

5   Contract Design Under Other-Regarding Preferences

The results from the gift exchange game surveyed previously indicate that reciprocity influences the relationship between principal and worker in that the principal can elicit more effort from the worker by offering a higher wage even though there are no guarantees that the worker will reciprocate by exerting a higher effort. This leads us to question how these deviations from the self-regarding model influence how best to structure the incentives that mediate the relationship between the two parties.

In this section we will apply what we have learned about other-regarding preferences to understand how they affect the optimal choice between competing types of contracts under the existence of moral hazard. More colloquially, we are interested in knowing whether it is best for the principal to reach for the carrot rather than the stick.

Consider a principal who contracts an agent to work for him. The agent can expend effort $e \in [\underline{e}, \bar{e}]$ at a cost $c(e)$ such that $c' > 0$ and $c'' > 0$, where $c(e)$ denotes the effort cost function. The principal wants the agent to expend e_{min}, which he introduces in the contract as the contracted effort level. However, e_{min} is non-binding since the agent's effort is not verifiable. Because the principal might want to provide evidence to a third party in the case where the worker is expending less effort than the contracted one the principal might invest in a verification technology that costs k. This technology is able to provide evidence of shirking with a probability $p \in (0, 1)$ which the principal can then use to impose a fine f on the agent.

In Fehr, Klein, and Schmidt (2007) a cohort of individuals are randomly selected into the roles of principal and agent and are then matched into principal-agent pairs. The game is played for 10 periods where in each one

a new pair is created so that no agent interacts more than once with the same principal. In the first stage of the game the principal chooses the type of contract and whether to invest in a verification technology; he proceeds to make an offer to the agent he is matched with. In the second stage the agent decides whether to accept the contract and, if he does, how much effort to expend. In the third and final stage, if the principal has invested in the verification technology, evidence of the agent's effort level is obtained with probability p and the terms of the contract are enforced.

The three types of contract the principal has at his disposal to offer are:

- **Incentive contract**: The contract specifies the wage w, the contracted effort level e_c, and the maximum fine $\bar{f}$ to be imposed on the agent if evidence of shirking is discovered. It is assumed the principal has invested in the verification technology.

- **Trust contract**: The contract specifies the wage w and the contracted effort level e_c but there is no investment in the verification technology. Therefore, e_c is non-verifiable and w is not contingent on the actual effort of the agent.

- **Bonus contract**: This contract is similar to the trust contract except where if the agent expends an effort level superior to e_c the principal might choose to reward the agent by offering a bonus b, which is not enforceable.

An analysis of what type of contract the principal would find preferable to offer under the assumption of self-regarding agents and principals is similar to the one made for the gift exchange game. Given the non-enforceability of e_c in both the trust and bonus contracts, and the added non-enforceability of b in the bonus contract, it is trivial to conclude that self-regarding agents will exert the lowest possible effort level.

However, the principal is able to induce a positive effort level from the agent if the verification technology is potent enough so that $p\bar{f} \geq c(e^*) - c(\underline{e})$. Therefore, under the assumption of self-regarding preferences, the principal will choose the incentive contract over the other two available alternatives: the trust and bonus contracts. This is a testable prediction.

Fehr et al. (2007) features two different experiments, one where principals can choose between a trust and an incentive contract and another where

principals choose between a bonus and an incentive contract. In the trust-incentive experiment the self-regarding preferences are largely confirmed. Most principals choose to offer agents an incentive contract and their share, over the 10 rounds played, increases over time. While there is experimentation on the part of most principals by offering a trust contract at least once, the differences in payoffs from both contracts were such that principals preferred the incentive contract. Once principals learned how to create an appropriate incentive contract with the right mix of w, e_c, and f, they made up the large majority of contracts offered.

Figure 6: Share of bonus and incentive contracts in the bonus-incentive experiment
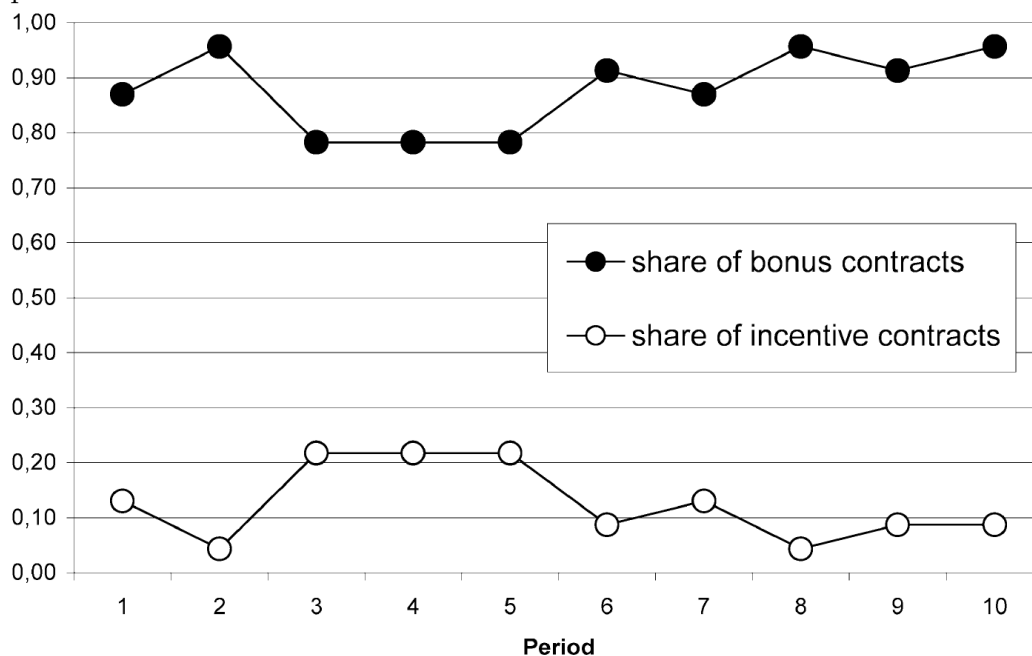
Source: Fehr et al. (2007)

It is in the bonus-incentive experiment that deviations from the self-regarding prediction are observed. Recall that the bonus b in the bonus contract is non-enforceable. If the principal, in his self-regarding rationale, decides not to offer a bonus, then the bonus contract becomes a trust one. Since agents are aware that the bonus is non-binding and therefore not likely to be realized, they should equate the bonus contract with the trust one and

29

act accordingly. This would lead us to predict that in the bonus-incentive experiment we should observe results similar to those from the trust-incentive one. This prediction is, however, not confirmed by the experimental results.

The overwhelming majority of contracts in the bonus-incentive experiment are bonus contracts, with the incentive contract seldom being offered. This choice is driven by the ability of bonus contracts to elicit a higher effort level from the agents which increases the payoff the principals get in comparison with the possible payoffs from the other contracts. Part of this larger surplus is then allocated by the principals as a bonus to the agents. The average income gained by agents in the bonus contract is approximately 23% higher than the one earned in the incentive contract. It turns out therefore that that the use of a bonus contract is beneficial to both parties.

Since the self-regarding model does not explain this set of choices, we will suppose an alternative where both the agent and the principal have other-regarding preferences of the Fehr-Schmidt form. This analysis follows closely Dhami (2016).

For simplicity we will assume that the agent's output, $\nu(e)$, is equal to e while the effort cost function is $c(e) = \frac{1}{2}e^2$ where $e \in [0,1]$. Under these specifications, the utility of a self-regarding agent and principal is given by, respectively

$$\begin{cases} u = & w - \frac{1}{2}e^2 - C_A \\ \pi = & e - w - C_p \end{cases} \tag{8}$$

where C_A and C_P are whatever other individual costs the agent and principal, respectively, incur by taking part in the relationship, such as the cost k of the verification technology for the principal. The respective FS preferences are given by

$$\begin{cases} U(u,\pi) = & u - \alpha_A \max\{\pi - u, 0\} - \beta_A \max\{u - \pi, 0\} \\ \Pi(u,\pi) = & \pi - \alpha_P \max\{u - \pi, 0\} - \beta_P \max\{\pi - u, 0\} \end{cases} \tag{9}$$

We make the additional assumption that α and β for both the agent and the principal is higher than 0.5, the reasoning being that under this assumption both parties will behave in ways such that their monetary payoffs are equalized. An indication of why this is so is provided in the Appendix. The parameter estimates gathered in Eckel and Gintis (2010) show more support for the assumption that $\alpha > 0.5$ than for $\beta > 0.5$, but the assumption affords us simplicity.

To show why the principal chooses to offer a bonus contract over the incentive contract we need to demonstrate that the former dominates the latter. We start by describing the expected profits for the principal under the incentive contract. The incentive compatibility constraint of the self-regarding agent is[4]

$$(1-p)\,w + p\,(w-f) \leq w - \frac{1}{2}e_c^2 \tag{10}$$

where we assume $C_A = 0$. This gives us $e_c \leq \sqrt{2pf}$ as the set of effort levels that are incentive compatible. To maximize profits, the self-interested principal sets a contract (w, e_c) that maximizes expected profits

$$E\,(\Pi) = (1-p)\,(e - w - k) + p\,(e - w - k + df) = e - w - k + pdf \tag{11}$$

where d is a binary variable dependent on whether $e < e_c$. If e_c satisfies the constraint that $e_c \leq \sqrt{2pf}$ and the constraint that $w > \frac{1}{2}e^2$ then $d = 0$. In light of these constraints we have that in an incentive contract in which the principal intends to maximize his profits the optimal contracted effort level is

$$e_I = \min\{1, \sqrt{2pf}\} \tag{12}$$

Recall that $e \in [0, 1]$. Equation (12) tells us the principal will choose to contract the minimum effort level that is also incentive compatible for the agent, which will depend on whether $\sqrt{2pf}$ is higher or lower than 1.

- If $\sqrt{2pf} \geq 1$ then we have $e_I = 1$ and $w = \frac{1}{2}$. In this case the expected profit of the firm is given by $E\,(\Pi) = \frac{1}{2} - k$.

- If $\sqrt{2pf} < 1$ then $e_I = \sqrt{2pf}$. Therefore $w = pf$ and $E\,(\Pi) = \sqrt{2pf} - pf - k$.

We now have the expected profits of the incentive contract which we can compare with the expected profit from a bonus contract offered by an other-regarding principal. The bonus contract is as described previously. In Stage 3, given that the experimental evidence suggests that $e > e_c$, the principal

---

[4]Under the incentive contract, both agent and principal act as self-regarding given that there is no opportunity for one to exhibit reciprocity towards the other.

awards a bonus b. Because he has other-regarding preferences, this bonus will be such that the payoffs of both parties are equaled.

Thus, we have

$$e - w - b = w + b - \frac{1}{2}e^2 \tag{13}$$

which we solve for b to get

$$b = \frac{1}{2}e - w + \frac{1}{4}e^2 = b(e, w) \tag{14}$$

The other-regarding agent will chose an optimal effort choice such that his monetary payoff is equal to that of the principal, that is,

$$w + b(e, w) - \frac{1}{2}e^2 = e - w - b(e, w) \tag{15}$$

Because the bonus is chosen so that the payoffs are equal, then (15) is satisfied for any value of e. Taking the first derivative of (15) in order to e gets us the result that the payoff is maximized at $e_F^b = 1$.

The other-regarding principal's expected payoff is $E(\Pi) = e - w - b(e, w)$, which when $e = 1$ yields

$$\begin{aligned} E(\Pi_B) &= e - w - \left(\frac{1}{2}e - w + \frac{1}{4}e^2\right) \\ &= 1 - w - \frac{1}{2} + w - \frac{1}{4} \\ &= \frac{1}{4} \end{aligned} \tag{16}$$

So given these options, which contract should the principal prefer? Under the self-regarding assumption we would expect the incentive contract to dominate over all others. However, taking into account that principals and agents might have other-regarding preferences, we conclude that the answer depends on a number of parameters.

Suppose that $\sqrt{2pf} > 1$, in which case $E(\Pi_I) = \frac{1}{2} - k$ as shown earlier. For the incentive contract to dominate over the bonus contract it would be needed that $E(\Pi_I) > E(\Pi_B)$, that is, $\frac{1}{2} - k > \frac{1}{4}$. This is only true if $k < \frac{1}{4}$. For the case where $\sqrt{2pf} < 1$ we have that $E(\Pi_I) = \sqrt{2pf} - pf - k$, which means that the incentive contract dominates over the bonus contract only if $\sqrt{2pf} - pdf - k > \frac{1}{4}$.

32

What these two situations show is that rather than the incentive contract always dominating over the bonus contract, it instead only does so when the deterrence parameters are high enough so that the principal has reliable access to evidence of shirking and the monitoring technology isn't too costly. If this isn't the case, because both parties have other-regarding preferences and go above and beyond their self-regarding behavior, the bonus contract engenders a relationship that is more beneficial to both the agent and the principal than the one created by the incentive contract.

Fehr and Schmidt (2007) extend the results from Fehr et al. (2007) by combining the fine from the incentive contract with bonus contract, creating a contract that features both the carrot and the stick. It was thought that given the combination of both incentives that this new contract would dominate over the others but that was not the case as more than two thirds of all contract offers were bonus contracts. The authors advance two possible explanations for their results. One is that the introduction of a fine might be seen by the agent as being in bad faith, leading them to reciprocate by expending a lower effort level. They also offer the hypothesis that since agents do not know the principal's trustworthiness they infer from the introduction of the fine that the principal will make a lower bonus offer. Indeed, from the experimental data, principals who offer the combined contract do offer significantly lower bonus payments.

6   Conclusion

It is our purpose as social scientists to venture farther into what we lack knowledge of and map out the intricacies that make up human behavior. We must observe the world around us, tease out its regularities, build up theories to explain them and test them against new observations. It is in the testing of those theories and the failure to explain behavior that we know our job is far from being over.

It has been argued throughout this study that sufficient experimental evidence has been accumulated to make us more doubtful about the assumption of agents possessing self-regarding preferences as sufficient to explain the full set of human behavior. This failure to explain documented behavior has motivated the introduction of the concept of other-regarding preferences, where agents are said to not only be preoccupied with themselves but also with those around them. We have introduced a new model that takes into

account other-regarding preferences and argued that this type of preferences allows us to explain what was previously an unexplainable behavior.

It is worth pointing out that this by no way means the self-regarding assumption is incorrect. That it has been continually used with success for many years shows that even though it is not a full description of how agents behave, it still is a useful modelling assumption of great explicability. Indeed, an issue that was not dealt with in this study is how to mediate between the two assumptions. Under what circumstances is one well served by the self-regarding assumption and under which should we introduce other-regarding preferences? The literature has thus far scantly addressed this issue and some guiding principles will need to emerge before more economists use these new preferences productively.

Regarding the experimental evidence used throughout this study, it is worth noting recent developments on the topic of replicability. Poor experimental procedure, ineffectual use of statistical tools, along with unwarranted confidence put on the results from the combination of the previous two being true, has led to the accumulation of false or irrelevant results. J. P. A. Ioannidis (2005) provides a good introduction to this problem. J. Ioannidis, Stanley, and Doucouliagos (2017) deals directly with Economics, where it is concluded that "nearly 80% of the reported effects [in the empirical economics literature surveyed in the paper] are exaggerated." Given the reliance in many of the experiments surveyed in this study on small sample sizes, and the resulting low statistical power, it would not be surprising if some of the conclusions they reach are not correct. While it was argued that laboratory evidence has a role in the study of economic behavior, this is only so if that evidence is properly gathered.

It is our hope that this study motivates the use of other-regarding preferences in the examination of economic behavior which either has not been sufficiently examined, or has only been so through the use of the self-regarding model. Economics would only gain by expanding the lens through which it studies behavior.

References

Akerlof, G. A. (1982). Labor contracts as partial gift exchange. *Quarterly Journal of Economics*, *97*(4), 543–569.

Andersen, S., Ertaç, S., Gneezy, U., Hoffman, M., & List, J. A. (2011). Stakes matter in ultimatum games. *The American Economic Review*, *101*(7), 3427–3439.

Andreoni, J. (1995). Cooperation in public-goods experiments: kindness or confusion? *The American Economic Review*, 891–904.

Barmettler, F., Fehr, E., & Zehnder, C. (2012). Big experimenter is watching you! anonymity and prosocial behavior in the laboratory. *Games and Economic Behavior*, *75*(1), 17–34.

Camerer, C. (2011a). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Camerer, C. (2011b). The promise and success of lab-field generalizability in experimental economics: A critical reply to Levitt and List. In G. R. Fréchette & A. Schotter (Eds.), *Handbook of experimental economic methodology*. Oxford University Press.

Camerer, C., Hogarth, R. M., Budescu, D. V., & Eckel, C. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. In *Elicitation of preferences* (pp. 7–48). Springer.

Cameron, L. A. (1999). Raising the stakes in the ultimatum game: Experimental evidence from indonesia. *Economic Inquiry*, *37*(1), 47–59.

Carter, J. R., & Irons, M. D. (1991). Are economists different, and if so, why? *Journal of Economic Perspectives*, *5*(2), 171–177.

Cohn, A., Fehr, E., & Goette, L. (2008). Fairness and effort-evidence from a field experiment. *Working Paper*.

Dana, J., Cain, D. M., & Dawes, R. M. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and Human Decision Processes*, *100*(2), 193–201.

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, *33*(1), 67–80.

Dhami, S. (2016). *The foundations of behavioral economic analysis*. Oxford University Press.

Ebbinghaus, H. (1885). *Memory: A contribution to experimental psychology*. New York: Dover.

Eckel, C., & Gintis, H. (2010). Blaming the messenger: Notes on the current state of experimental economics. *Journal of Economic Behavior and Organization*, *73*(1), 109–119.

Edgeworth, F. Y. (1881). *Mathematical psychics: An essay on the application of mathematics to the moral sciences* (Vol. 10). Kegan Paul.

Falk, A., & Heckman, J. J. (2009). Lab Experiments Are a Major Source of Knowledge in the Social Sciences. *Science*, *326*(5952), 535–538.

Falk, A., Meier, S., & Zehnder, C. (2013). Do lab experiments misrepresent social preferences? the case of self-selected student samples. *Journal of the European Economic Association*, *11*(4), 839–852.

Fehr, E., & Falk, A. (2002). Psychological foundations of incentives. *European Economic Review*, *46*(4-5), 687–724.

Fehr, E., Gächter, S., & Kirchsteiger, G. (1997). Reciprocity as a contract enforcement device: Experimental evidence. *Econometrica*, 833–860.

Fehr, E., Goette, L., & Zehnder, C. (2009). A Behavioral Account of the Labor Market: The Role of Fairness Concerns. *Annual Review of Economics*, *1*(1), 355–384.

Fehr, E., Kirchsteiger, G., & Riedl, A. (1993). Does fairness prevent market clearing? an experimental investigation. *Quarterly Journal of Economics*, *108*(2), 437–459.

Fehr, E., Klein, A., & Schmidt, K. M. (2007). Fairness and contract design. *Econometrica*, *75*(1), 121–154.

Fehr, E., & List, J. A. (2004). The hidden costs and returns of incentives—trust and trustworthiness among CEOs. *Journal of the European Economic Association*, *2*(5), 743–771.

Fehr, E., & Schmidt, K. (1999). A Theory of Fairness,Competition and Cooperation. *Quarterly journal of Economics*, *114*(August), 817–868.

Fehr, E., & Schmidt, K. M. (2007). Adding a stick to the carrot? The interaction of bonuses and fines. *American Economic Review*, *97*(2), 177–181.

Gneezy, U., & List, J. A. (2006). Putting behavioral economics to work: resting for gift exchange in labor markets using field experiments. *Econometrica*, *74*(5), 1365–1384. doi: 10.1111/j.1468-0262.2006.00707.x

Guala, F. (2008). History of Experimental Economics. In *The new palgrave dictionary of economics* (pp. 1–8). Plagrae Macmillan.

Hausman, D. M. (2011). *Preference, value, choice, and welfare.* Cambridge University Press.

Henrich, J. (2000). Does culture matter in economic behavior ? Ultimatum

Game Bargaining among the Machiguenga of the Peruvian Amazon. *American Economic Review*, *4* (October).

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., ... Others (2005). "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, *28* (6), 795–815.

Homans, G. C. (1954). The cash posters: a study of a group of working girls. *American Sociological Review*, *19* (6), 724–733.

Ioannidis, J., Stanley, T. D., & Doucouliagos, H. (2017). The power of bias in economics research. *The Economic Journal*, *127* (605).

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS medicine*, *2* (8), e124.

Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, *32* (5), 865–889. doi: 10.1016/j.joep .2011.05.007

Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, *21* (2), 153–174.

List, J. A. (2004). Young, selfish and male: Field evidence of social preferences. *Economic Journal*, *114* (492), 121–149.

List, J. A. (2006). The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions. *Journal of Political Economy*, *114* (1), 1–37. doi: 10.1086/498587

List, J. A. (2007). On the interpretation of giving in dictator games. *Journal of Political economy*, *115* (3), 482–493.

Mearsheimer, J. J., & Walt, S. M. (2013). Leaving theory behind: Why simplistic hypothesis testing is bad for international relations. *European Journal of International Relations*, *19* (3), 427–457.

Moscati, I. (2007). Early experiments in consumer demand theory: 1930-1970. *History of political economy*, *39* (3), 359.

Roth, A. E., Prasnikar, V., Okuno-Fujiwara, M., & Zamir, S. (1991). Bargaining and market behavior in jerusalem, ljubljana, pittsburgh, and tokyo: An experimental study. *The American Economic Review*, 1068–1095.

Samuelson, P., & Nordhaus, W. D. (1985). *Economics*. McGraw-Hill.

Samuelson, P., & Nordhaus, W. D. (1992). *Economics*. McGraw-Hill.

Slonim, R., & Roth, A. E. (1998). Learning in high stakes ultimatum games: An experiment in the slovak republic. *Econometrica*, 569–596.

Smith, A. (1822). *The theory of moral sentiments* (Vol. 1). J. Richardson.

Ward, A. (1997). Naive realism in everyday life: Implications for social conflict and misunderstanding. *Values and knowledge*, 103–135.

7 Appendix

7.1 Contract Design Under Other-Regarding Preferences (p.31)

Consider the agent's problem of choosing an effort level. We intend to show that an other-regarding agent with $\alpha_A \geq \beta_A > 0.5$ will choose e such that his monetary payoff is equal to the principal's.

- $\mathbf{x_A > x_P}$

If the agent's monetary payoff is higher than the principal's then his utility is $U(x_A, x_P) = x_A - \beta_A(x_A - x_P)$. Suppose the agent transfers $\epsilon > 0$ to the principal. Then,

$$U(x_A, x_P) = x_A - \epsilon - \beta\left[(x_A - \epsilon) - (x_P + \epsilon)\right]$$
$$= x_a - \beta(x_A - x_P) + \epsilon(2\beta_A - 1)$$

which implies a positive change since $\beta_A > 0.5$. The agent is therefore made better off by transferring resources to the principal.

- $\mathbf{x_P > x_A}$

Now consider the case where the agent's monetary payoff his lower than the principal's. The agent can punish the principal and reduce his payoff by a unit at cost $\gamma < 1$. The agent's utility if he does so is

$$U(x_A, x_P) = (x_A - \gamma) - \alpha_A\left[(x_P - 1) - (x_A - \gamma)\right]$$
$$= x_A - \alpha_A(x_P - x_A) + \alpha_A(1 - \gamma) - \gamma$$

For the change to be positive, we need that the last term in the right hand side of the equation to be positive. This means that $\gamma < \frac{\alpha_A}{1+\alpha_A} < 1$.

A similar argument can be made for how the principal reacts to inequity, leading us to conclude that if both have $\alpha \geq \beta > 0.5$ they will work towards equaling their monetary payoffs.