Grohn, Jan; Huck, Steffen; Valasek, Justin Mattias

Working Paper

A note on empathy in games

WZB Discussion Paper, No. SP II 2014-301

Provided in Cooperation with:
WZB Berlin Social Science Center

Suggested Citation: Grohn, Jan; Huck, Steffen; Valasek, Justin Mattias (2014) : A note on empathy in games, WZB Discussion Paper, No. SP II 2014-301, Wissenschaftszentrum Berlin für Sozialforschung (WZB), Berlin

This Version is available at:
https://hdl.handle.net/10419/91719

WZB

Wissenschaftszentrum Berlin
für Sozialforschung

Jan Grohn
Steffen Huck
Justin Mattias Valasek

A Note on Empathy in Games

Discussion Paper

SP II 2014–301

January 2014

Research Area
Markets and Choice

Research Unit
Economics of Change

Wissenschaftszentrum Berlin für Sozialforschung gGmbH
Reichpietschufer 50
10785 Berlin
Germany
www.wzb.eu

Affiliation of the authors:

Jan Grohn, University of Bayreuth
Steffen Huck, WZB Berlin and University College London
Justin Mattias Valasek, WZB Berlin

Abstract

A Note on Empathy in Games

In this note we shall discuss a concept that – despite its prominence in both Hume (1739) and Smith (1759), its obvious relevance for social behavior, and its not so infrequent use in colloquial language – has never gained a foothold in economic theory: the concept of empathy. Specifically, we illustrate how some insights from the psychological literature on empathy can be incorporated into a standard utility framework, and demonstrate the potential interaction of beliefs and utility through the channel of empathy.

In diesem Artikel diskutieren wir das Konzept der Empathie. Dieses konnte in der ökonomischen Theorie nie wirklich Fuß fassen, trotz seiner Bedeutung sowohl bei Hume (1739) als auch Smith (1759), seiner offensichtlichen Relevanz für soziales Verhalten und seines durchaus verbreiteten Gebrauchs in der Umgangssprache. Insbesondere zeigen wir, wie einige Erkenntnisse aus der psychologischen Literatur über Empathie in ein Standardkonzept von ‚Nutzen' integriert werden können und demonstrieren die potenzielle Interaktion von Erwartungen und Nutzen über den Weg der Empathie.

Keywords: Empathy, Belief Formation, Preferences.

JEL classification: D03, D83.

1 Introduction

Ever since the publication of Werner Güth's ultimatum game experiment in this journal (Güth, Schmittberger, and Schwarze 1982) economists have been on a quest to incorporate social concerns into their theories and, by now, there is an abundance of models capturing ideas like inequality aversion (Bolton 1997; Fehr and Schmidt 1999; Bolton and Ockenfels 2000), fairness and reciprocity (Rabin 1993; Dufwenberg and Kirchsteiger 2004; Falk and Fischbacher 2006; Cox, Friedman and Gjerstad 2007), efficiency concerns (Charness and Rabin 2002), or impure altruism (Andreoni 1990). These models have proven attractive to economists as they can help to organize data and are at the same time familiar: agents are endowed with a utility function, form (correct) beliefs about others and play optimally given these beliefs. Nash equilibrium remains the core solution concept for all these models (appropriately adapted to allow beliefs to enter utility functions directly as in Geanakoplos, Pearce and Stacchetti 1989 and Battigalli and Dufwenberg 2009). In his 1995 survey (again in this journal) Güth branded these approaches as being part of a 'neoclassical repair shop.' This repair shop would turn into a giant industry reshaping neoclassical economics itself.

As an alternative, Güth has been advocating the exploration of more psychological, process-orientated models as a more promising agenda that abandons the as-if maximization doctrine of the orthodoxy (Güth 2000; Güth and Kliemt 2004; Güth 2013).

In this note we shall discuss a concept that – despite its prominence in both Hume (1739) and Smith (1759), its obvious relevance for social behavior, and its not so infrequent use in colloquial language – has never gained a foothold in economic theory: the concept of empathy.

The term empathy is prominent in both the psychology and philosophy literature, and can mean many different things. In this note we will focus on the key psychological theories of empathy and explore their relevance for choice behavior. We will demonstrate how empathy is set apart from other popular concepts in behavioral economics, which typically fall in one of two categories: "biases" affecting belief formation, and "biases" affecting utility. Empathy, we will show, is operating through both belief and utility formation. Hence, in the presence of empathy, beliefs and utility become intricately linked – indeed manipulations and even mere measurement of beliefs can change valuations/utility.

In this note we offer a brief discussion of some key notions of empathy taken from the psychology literature and introduce the so-called "empathy-altruism hypothesis" that suggests that a particular process of belief formation triggers partial altruism. We will suggest a simple model to capture this link and apply it to three toy games highlighting the subtle nature of empathy. While our model sketch satisfies Güth's demand for more process-oriented psychological modeling it does not break with the convention of as-if maximization. Finally, we offer some concluding thoughts and suggest areas of experimental economics that could be re-assessed in the light of the empathy-altruism hypothesis.

2 Imagine-self, imagine-other, and the empathy-altruism hypothesis

Batson (2009) and Stueber (2013) discuss various aspects and definitions of empathy, ranging from simple knowledge of others' emotions to directly experiencing the emotions of others. In this note we focus on two aspects of empathy. We first introduce two mechanisms, imagine-self and imagine-other, that define empathy as a process of belief formation. Subsequently, we sketch Batson's (2011) empathy-altruism hypothesis, which proposes a link between belief formation and valuations.

The first aspect of empathy refers to people's capability to infer what others think or feel. This is referred to as "mind reading" by Batson (2009). Mind reading is in itself not an emotional process, but is of obvious importance for belief formation in strategic contexts. It is exactly what a player has to do when predicting another player's choice (which includes, of course, inferences about the other's beliefs). But what is the process by which an individual actually forms these inferences? There are two processes commonly considered in the psychology literature that we will integrate into our model: "imagine self" and "imagine other."

With "imagine self" people try to imagine themselves in other people's shoes, that is, they try to imagine their own emotions in similar circumstances. "Imagine self" is related to the often documented false-consensus bias where subjects tend to believe that others are similar to them (see, e.g., Ross, Greene, and House 1977). An alternative form of perspective taking is "imagine other" where a person tries to imagine how another person is feeling. While "imagine other" ostensibly requires a greater degree of sophistication, these are both high-level processes, related to Smith's (1759) ideas of mindreading.[1]

The second aspect of empathy considers whether the process of inferring others' emotions has any consequences on the individual's valuations. It is possible that this process has no consequences at all. We might perfectly understand the misery of a person and yet ignore it. However, the psychology literature on empathy suggests otherwise; in particular, the literature suggests strong links between imagine-other and altruistic behavior, a link that is not present when individuals use imagine-self. In summarizing three decades of research, Batson (2011) claims that imagining others triggers altruistic motivations, that is, the ultimate desire to help another person. In principle, Batson argues people might also help others (i) to reduce personal distress, (ii) to avoid social sanctions, and (iii) to gain social rewards. He shows how helping remains a stable phenomenon even if these three channels are shut down, supporting the empathy-altruism hypothesis. However, not everybody helps in these experiments. Batson shows that the key route to generate altruism and thus trigger helping is to trigger "imagine other." While the experiments in this literature are far-removed from the style of economic

[1] In contrast, Hume (1739) had stressed the automatic character of assessing others, more closely related to concepts of mimicry and mirroring, which we do not discuss here. See the introduction to Coplan and Goldi (2011) for more about these ancestors of theories of empathy.

experiments – they typically involve elaborate stories about people in need – there are interesting parallels to the ultimatum-dictator game debate. Hoffman, McCabe, and Smith (1996), for example, varied social distance between dictators and recipients and showed how giving declined with distance – a finding that potentially could be reinterpreted in light of Batson's hypothesis.

Below we will develop some toy games to explore the theoretical predictions of the empathy-altruism hypothesis in a strategic context. In order to test the predictions we will derive, it would be possible to follow Batson and employ different forms of priming in order to manipulate the extent to which subjects employ "imagine other" instead of "imagine self."

2.1 Modeling the empathy-altruism hypothesis and three toy games

Next we introduce a simple framework to illustrate the implications of the empathy-altruism hypothesis for choice. Specifically, following Batson (2011) we explore the possibility that agents who are more sophisticated when it comes to evaluating the preferences of others are also more prone to have "other-regarding" preferences. Reasonably, when agents accurately observe the emotional state of others, they might be more prone to be influenced by others' emotions. The reverse causality might also play a role: if an agent places value on the welfare of others, they might invest more cognitive power on evaluating the preferences, and hence welfare, of others.

This suggests an interesting correlation between beliefs and utility. Agents who use "imagine other" are more likely to care about others, suggesting that altruistic agents have more accurate beliefs than purely selfish agents. In other words, a model capturing the empathy-altruism hypothesis will not nest homo oeconomicus – accurate beliefs and selfishness do not go hand in hand according to the hypothesis. This correlation might confound experimental findings. Agents with other-regarding preferences have more accurate beliefs of other agents' types, and are therefore "better" strategic players. Below we illustrate this in the context of three toy games, a public good game, an ultimatum game and a battle of the sexes game. As we will show, empathy can interact with choice in a rather counterintuitive manner.

FRAMEWORK: We consider agents that care about both their total and relative levels of consumption.[2] Specifically, all agents have utility functions over consumption, c_i and c_j, of the following form:

$$u_i(c_i, c_j) = c_i + v(c_i - c_j),$$

where for $x > 0$, $v(x) > 0$, $v'(\cdot) > 0$, and $v''(x) < 0$; also, $v(\cdot)$ is symmetric in the sense that $v(x) = -v(-x)$.[3]

[2] We consider agents with preferences over relative consumption due to the ultimatum game we consider below; only agents with some preference over relative income will reject offers in the ultimatum game.

[3] For convenience, we use a model in which agents receive positive payoffs for positive relative levels of income, and equivalent negative payoffs for negative relative levels of income. We discuss the robustness of our results to a model in which agents only receive negative payoffs for negative relative income;$v(c_i, c_j) = v(min\{(c_i - c_j), 0\})$, i.e. Fehr-Schmidt (1999) inequality aversion with $\beta = 0$.

Additionally, agents are one of two types: a low-empathy type (type L), and a high-empathy type (type H). High empathy, whether inherent or primed, affects agents through two channels: (1) they are more likely to directly value the wellbeing of other agents, (2) they are more likely to form sophisticated beliefs regarding the preferences of others.

Through the first channel, the high-empathy type, H, generates *altruism* in the sense that they partially internalize the well-being of the other agent. We model this as partial altruism within the standard utility-maximization framework. That is, i of type H has overall preferences represented by the utility function:

$$w_i(c_i, c_j) = u_i(c_i, c_j) + \lambda u_j(c_i, c_j),$$

where $\lambda \in (0, 1)$. The low-empathy type, on the other hand, simply has preferences represented by the utility function $w_i(c_i, c_j) = u_i(c_i, c_j)$.

Through the second channel, the high-empathy type employes a more sophisticated method of evaluating the types of others, relative to the low-empathy type. Specifically, we assume that agents of type H employ "imagine other," while agents of type L employ "imagine self." For simplicity, assume that "imagine self" results in the belief that other agents are also type L, while "imagine other" results in accurate beliefs regarding the other's type.[4]

Having established types, we now model their behavior in a set of illustrative games.

GAME 1: Pairs of agents, i and j, interact in a public goods contribution game, where each agent is endowed with income y, and income can be contributed to a "common pool." Contributions, x_i and x_j, are multiplied by a factor of $\delta \in (1, 2)$ and divided equally between the two players.

In this public goods game type L, who is not altruistic, will not contribute. As is seen in the following equation, choosing zero contributions, $x_i = 0$, maximizes c_i regardless of the actions of the other player.

$$c_i(x_i, x_j) = y + \frac{\delta}{2} x_j + \left(\frac{\delta}{2} - 1\right) x_i.$$

Therefore, independent of the other player's type, L maximizes both individual and relative income $(c_i - c_j)$ by not contributing.

Since L sets $x_i = 0$, any positive contributions in the public goods game must come from type H. Type H might contribute, since she directly values the utility of the other player, and since contributions "produce" higher levels of aggregate consumption. Specifically, type H will contribute if λ and δ are large enough to counter the loss of individual and relative income from setting $x_i > 0$ (see the following equation):

$$w_i(x_i, x_j) = (1 + \lambda)y + \left(\frac{\delta}{2} - \lambda\right) x_j + \left((1 + \lambda)\frac{\delta}{2} - 1\right) x_i + v(c_i, c_j) + \lambda v(c_j, c_i).$$

[4]The examples below give analogous results if H instead has accurate beliefs of the composition of preferences in the experimental pool.

The level of contribution type H chooses is a function of the type of the other player (player j). This conditionality does not stem from H valuing j's utility more if she is also type H; H is equally altruistic towards players of type L and H. Instead, this relationship arises because contributions are strategic complements. Contributions by player j increase i's relative and total income, which decreases the relative cost of x_i. Therefore, it is easy to see that in any equilibrium between two players of type H, i's contribution will be weakly higher than if player j is of type L.[5]

To summarize, Game 1 gives the following predictions: (1) Only players of type H will contribute to the common pool, and (2) players of type H will contribute more if they are paired with another player of type H. The predictions of the public goods contribution game conform to the intuitive characteristics of an "other-regarding" or altruistic behavioral type. In this game, players that are empathetic also take an action that might be characterized as "altruistic" and also exhibit reciprocal behavior.

GAME 2: Pairs of agents, i and j, interact in an ultimatum game, where i proposes a split of y dollars, $\{c_i, c_j\}$, and j chooses whether to accept or reject. If j accepts, y is allocated to the players according to the proposed split (i receives c_i, j receives c_j); if j rejects, both players receive a payoff of zero.

For a non-empathetic player i, Game 2 is more 'complex' than Game 1 due to the strategic interaction with player j: in Game 2, player i must account for the possibility that player j conditions her action on the proposed split. Specifically, since j values both total and relative income, j will reject a split that results in a high level of inequality. Therefore, player i of type L will choose the split with the highest value of c_i that will not be rejected by player j. That is, i will choose the split that solves the following problem:

$$max_{[c_i,c_j]} w_i(c_i, c_j), \text{ subject to} \tag{2.1}$$

$$w_j(c_i, c_j) \geq 0 \tag{2.2}$$

The highest level of inequality that j is willing to accept depends on j's type. If j is of type L, then the highest level of inequality j is willing to accept is the split where the negative utility from the relative income differential, $v(c_j - c_i)$, is precisely offset by the positive utility from consumption, c_j. Label this split, where $v(c_j - c_i) = c_j$, $\{c_i^l, c_j^l\}$. If j is of type H, however, then j is willing to accept a higher level of inequality than $\{c_i^l, c_j^l\}$. This is because players of type H also receive positive utility from the consumption of the other player. Therefore, the highest level of inequality j of type H is willing to accept is the split where $v(c_j - c_i)$ is offset by the *sum of* c_j *and* $\lambda\omega_i(c_i, c_j)$. This split,

[5] Contributions are also strategic complements under inequality aversion ($v(c_i, c_j) = v(min\{(c_i - c_j), 0\})$); since players of type H are adverse to inequality regardless of whether their relative income is negative (direct inequality aversion) or positive (indirect inequality aversion due to altruism), higher contributions by player j decrease the inequality cost of contributing (or increase the benefit).

labeled $\{c_i^h, c_j^h\}$, allocates strictly more to player i than $\{c_i^l, c_j^l\}$.[6]

Returning to i's choice of which split to offer, a non-empathetic, and hence non-sophisticated, i holds the belief that player j is also of type L. Therefore, even though i seeks to maximize c_i, she will offer j the more equal split $\{c_i^l, c_j^l\}$ regardless of j's type.

Again, an empathetic player i differs from the non-empathetic type in that she is more sophisticated in forming beliefs of j's type, and also values the utility of player j. Note that since H is only "partially altruistic" ($\lambda \in [0, 1)$), her overall preference function is strictly increasing in $(c_i - c_j)$, given the constraint $c_i + c_j = y$; that is, since i values the other player's consumption at a comparatively lower rate than her own, she always prefers a one-to-one trade of c_j for c_i.[7] This implies that, in a game such as the ultimatum game where the surplus is constant (assuming the proposed split is accepted), type H will choose the split that maximizes her own consumption. Moreover, since H utilizes "imagine other" and is sophisticated in her beliefs regarding the type of the other player, she always offers the minimum split that j is willing to accept: an i of type H will offer $\{c_i^l, c_j^l\}$ to a j of type L, and offer $\{c_i^h, c_j^h\}$ to a j of type H. Therefore, in the ultimatum game, the empathetic type will on average offer less than the non-empathetic type.

The analysis of the two games shows that, relative to type L, type H will take actions that are more favorable to the other player in Game 1, and actions that *are less favorable to the other player in Game 2*. This surprising and perhaps counter-intuitive result occurs precisely because empathy affects action through two channels: (1) sophisticated beliefs regarding the other's type, and (2) valuing the well-being of the other. In Game 1, the second channel dominates; since the amount of surplus can be increased, there is an incentive for empathetic agents to act "prosocial." In Game 2, the first channel dominates; since surplus is constant in case of agreement, the incentive to "share" is driven solely by the possibility of rejection by the other agent. Here, agents with sophisticated beliefs will utilize this information to maximize their own material payoff.

An empirical regularity of experimental economics is that some subjects behave in a "prosocial" manner. A researcher might conclude that prosocial behavior is simply the result of a prosocial type. Instead, as we illustrate here, seeming differences in "prosocial" behavior might instead stem from differences in the game structure, interacted with different degrees of empathy.

GAME 3: Lastly, we consider the implications of the empathy-altruism hypothesis for the battle of the sexes. Specifically, we consider the mixed strategy equilibrium for the game in Figure 1, which generally provides a good fit for aggregate play in an experimental setting.

First, consider the play of two non-empathetic agents. Both use imagine-self, and therefore hold

[6] The minimum split that j of type H is willing to accept depends on i's type; take $\{c_i^h, c_j^h\}$ to be the minimal split when i is of type H. Note, that even if we used the alternative definition, $c_i^h > c_i^l$.

[7] This prediction holds for all parameter values due to the assumed symmetry of $v(c_i - c_j)$. With $v(c_i, c_j) = v(min\{(c_i - c_j), 0\})$, $\omega_i(c_i, c_j)$ will be decreasing in c_i for high values of λ and c_i. As long as $\omega_i(c_i, c_j)$ is increasing in c_i at $\{c_i^l, c_j^l\}$, which is true as long as λ is not too high, then our results carry through with inequality aversion.

	B	S
B	$(1, \alpha)$	(0,0)
S	(0,0)	$(\alpha, 1)$

Figure 1: Bach versus Stravinsky (utility payoffs; $\alpha \in (0, 1)$).

correct beliefs regarding the type of the other player. In equilibrium, both agents will play their preferred option with probability $\sigma^{ll} = 1/(1+\alpha)$, implying that coordination will be successful with probability $2\sigma^{ll}(1-\sigma^{ll})$.

Second, consider the play of two empathetic types. Both hold correct beliefs regarding the other player. However, compared to type L players, type H players have a lower utility difference between coordination on their preferred and less preferred option, since partial altruism increases the payoff of (B, B) by $\lambda\alpha$ and (S, S) by λ. Therefore, in equilibrium, both agents will play their preferred option with probability $\sigma^{hh} < \sigma^{ll}$. This implies the intuitive result that two empathetic players will coordinate more often than two non-empathetic players (i.e. $2\sigma^{hh}(1-\sigma^{hh}) > 2\sigma^{ll}(1-\sigma^{ll})$).

Next, consider the play of an empathetic type matched with a non-empathetic type. Assume for convenience that the player of type L prefers Stravinsky. In this case, type L has incorrect beliefs regarding the type of the other player, while type H has correct beliefs (we also assume that type H has correct higher-order beliefs). Therefore, type L will continue to play S with probability σ^{ll}. Type H, however, will now play B with zero probability. To see why this is H's best response, note that H is indifferent between B and S when the other player plays S with probability σ^{hh}; since a plays S with probability $\sigma^{ll} > \sigma^{hh}$, type H strictly prefers to play S over B.

Note that this implies that coordination will occur with higher probability (σ^{ll}) between types L and H when compared to coordination with two L types ($2\sigma^{ll}(1-\sigma^{ll})$). Less intuitively, however, expected coordination between types L and H is also *higher than between two H types.*[8] This higher rate of coordination is a result of type L's use of "imagine self," which functions as a credible commitment device to play her preferred option at a higher-than-equilibrium rate, causing a player of type H to yield and play her less-preferred option.

The predictions of Game 3 provide a potential strategy for testing the empathy-altruism hypothesis. Assume a treatment mechanism for increasing the frequency of the empathetic type (or, equivalently,

[8] Note that $2\sigma^{ll} > 2\sigma^{hh}$ or, equivalently $\frac{\sigma^{ll}}{1/2} > 2\sigma^{hh}$. Also, since $(1-\sigma^{hh}) < \frac{1}{2}$, this implies that $\frac{\sigma^{ll}}{(1-\sigma^{hh})} > 2\sigma^{hh}$, or $\sigma^{ll} > 2\sigma^{hh}(1-\sigma^{hh})$.

triggering higher levels of empathy) exists; e.g. encouraging the subject to consider the other player by eliciting beliefs regarding the other player's actions, or pro-social framing. The empathy-altruism hypothesis predicts that subjecting all players to the treatment will increase coordination. This prediction, however, is also in line with a simple increase in altruism. However, if the treatment is applied asymmetrically, i.e. only one player in each pair is subjected to the treatment, *and* the treated players are informed that the other player did not receive the treatment, then the empathy-altruism hypothesis predicts coordination will increase even more than when both players are treated.

3 Concluding thoughts

In economic models of other-regarding preferences, preferences over the others' consumption, such as altruism, are often modeled as if individuals know, and directly value, the utility of others. This approach has the advantage of being a parsimonious deviation from the utility model over bundles of own consumption. However, even a small step away from traditional economic preferences implies a large leap in psychological complexity. Assuming an individual prefers more consumption to less is a simple assumption to make; characterizing how they assess and value the well-being of others is a much more complex task. In economics, much has been done regarding how individuals value the well-being of others. How they assess others' wellbeing, however, has received comparatively little attention in the economic literature.

In this note, we present a simple economic model capturing the empathy-altruism hypothesis, a prominent result in psychology, and apply the model to three simple toy games that shed some light on the potential complexity of the interaction of empathy and choice. Crucially, the empathy-altruism hypothesis entails that agents with partially altruistic preferences tend to have more accurate beliefs about others. There is now a rapidly growing literature on belief elicitation, and one of the major concerns is whether methods of belief elicitation affect how subjects play the game (see, e.g., Rustrom and Wilcox 2009). Viewing this literature from the empathy angle, the question is, of course, whether there is a "natural" way of playing a game. What would that be? The behavior observed in another treatment without belief elicitation? Why? Maybe the idea of searching for "neutral" elicitation methods is misguided and we can learn something important about mindreading from elicitation procedures that do affect game play.

In experimental economics, instructions may have a strong influence on how subjects assess what others think or feel and we have alluded to the role of priming for possible experimental investigations into the role of empathy. In the mid-nineties Güth initiated a research program on "decision preparation" exploring different ways of aiding subjects' deliberations before playing a game. For example, Güth, Huck, and Ockenfels (1996) explores treatments where subjects had to answer very detailed pre-experimental questionnaires about the three-player ultimatum game with incomplete information that they were about to play. The questions were designed to encourage backward induction and encour-

aged both "imagine other" and "imagine self." Remarkably, the questionnaire led proposers to more generosity. In the light of our analysis above, it would be interesting to re-run a similar experiment with two different questionnaires, one fostering "imagine other," the other fostering "imagine self."

There are various other findings in the experimental literature that could be re-assessed in light of the empathy-altruism hypothesis. For example, in many trust experiments one finds that first movers react much more sluggish to changes in the environment than second movers (see, for example, Bohnet et al. 2005). Interestingly, this is in line with the framework that we propose above. Suppose that there are two trust games, A and B, such that trustees with partial altruism will tend to reciprocate trust in variant B more than in variant A due to slight differences in the payoff structure. Low-empathy trusters will play both games in the same way: they will imagine that second movers are not trustworthy (as they would not reciprocate trust themselves), and hence will never trust. At the same time, some partially altruistic trusters will already trust in game A as they partially internalize the higher efficiency generated through trust even if they expect to be exploited. Thanks to the link between altruism and beliefs, as the game shifts from A to B, trusters will not react as sharply to changes in trustworthiness as models with rational expectations would predict.

Finally, a recent study by Falk and Szech (2013) demonstrates how subjects' willingness to pay for saving the life of a mouse drops in market environments. While they attribute their finding to corrosion of moral values, an alternative hypothesis is that the market changes subject's perception of the mouse's suffering when gassed to death. Negotiating a market environment might reduce the amount of attention that subjects can devote to "imagine the mouse," and others' willingness to accept the death of the mouse might completely block this process.

References

[1] James Andreoni. Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. *The Economic Journal*, 100(401):464–477, June 1990.

[2] Daniel Batson. These things called empathy: Eight related but distinct phenomena. In Jean Decety and William Ickes, editors, *The social neuroscience of empathy*, pages 3–15. MIT Press, Cambridge, 2009.

[3] Daniel Batson. *Altruism in Humans*. Oxford University Press, Oxford, 2011.

[4] Pierpaolo Battigalli and Martin Dufwenberg. Dynamic psychological games. *Journal of Economic Theory*, 144(1):1–35, January 2009.

[5] Iris Bohnet, Heike Harmgart, Steffen Huck, and Jean-Robert Tyran. Learning Trust. *Journal of the European Economic Association*, 3(2-3):322–329, April 2005.

[6] Gary Bolton. The rationality of splitting equally. *Journal of Economic Behavior & Organization*, 32(3):365–381, March 1997.

[7] Gary Bolton and Axel Ockenfels. ERC: A Theory of Equity, Reciprocity, and Competition. *The American Economic Review*, 90(1):166–193, March 2000.

[8] Gary Charness and Matthew Rabin. Understanding Social Preferences with Simple Tests. *The Quarterly Journal of Economics*, 117(3):817–869, 2002.

[9] Amy Coplan and Peter Goldie. *Empathy: Philosophical and Psychological Perspectives*. Oxford University Press, Oxford, 2011.

[10] James Cox, Daniel Friedman, and Steven Gjerstad. A tractable model of reciprocity and fairness. *Games and Economic Behavior*, 59:17–45, 2007.

[11] Martin Dufwenberg and Georg Kirchsteiger. A theory of sequential reciprocity. *Games and Economic Behavior*, 47(2):268–298, 2004.

[12] Armin Falk and Urs Fischbacher. A theory of reciprocity. *Games and Economic Behavior*, 54(2):293–315, February 2006.

[13] Armin Falk and Nora Szech. Morals and Markets. *science*, 340(6133):707–711, 2013.

[14] Ernst Fehr and Klaus Schmidt. A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics*, 113(3):817–868, August 1999.

[15] John Geanakoplos, David Pearce, and Ennio Stacchetti. Psychological games and sequential rationality. *Games and Economic Behavior*, 1(1):60–79, March 1989.

[16] Werner Güth. On ultimatum bargaining experiments—A personal review. *Journal of Economic Behavior & Organization*, 27(3):329–344, 1995.

[17] Werner Güth. How ultimatum offers emerge: A study in bounded rationality. *Mimeo*, 2000.

[18] Werner Güth. Observing Mental Modeling. *Review of Behavioral Economics*, (Forthcoming), 2013.

[19] Werner Güth, Steffen Huck, and Peter Ockenfels. Two-Level Ultimatum Bargaining with Incomplete Information: an Experimental Study. *The Economic Journal*, 106(436):593–604, May 1996.

[20] Werner Güth and Hartmut Kliemt. Bounded rationality and theory absorption. *Mimeo*, 2004.

[21] Werner Güth, Rolf Schmittberger, and Bernd Schwarze. An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, 3(4):367–388, December 1982.

[22] Elizabeth Hoffman, Kevin McCabe, and Vernon L Smith. Social Distance and Other-Regarding Behavior in Dictator Games. *The American Economic Review*, 86(3):653–660, June 1996.

[23] David Hume. *A Treatise of Human Nature*. Oxford University Press, Oxford, 1739.

[24] Matthew Rabin. Incorporating Fairness into Game Theory and Economics. *The American Economic Review*, 83(5):1281–1302, December 1993.

[25] Lee Ross, David Greene, and Pamela House. The "false consensus effect:" An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, 13(3):279–301, May 1977.

[26] Elisabet Rutström and Nathaniel Wilcox. Stated beliefs versus inferred beliefs: A methodological inquiry and experimental test. *Games and Economic Behavior*, 67(2):616–632, November 2009.

[27] Adam Smith. The Theory of Moral Sentiments. *Millar*, 1759.

[28] Karsten Stueber. The Causal Autonomy of Reason Explanations and How Not to Worry about Causal Deviance. *Philosophy of the Social Sciences*, 43(1):24–45, January 2013.

Discussion Papers of the Research Area Markets and Choice 2013

Research Unit: **Market Behavior**

Nadja Dwenger, Dorothea Kübler, Georg Weizsäcker SP II 2013-201
Preference for Randomization: Empirical and Experimental Evidence

Kai A. Konrad, Thomas R. Cusack SP II 2013-202
Hanging Together or Being Hung Separately: The Strategic Power of Coalitions where Bargaining Occurs with Incomplete Information

David Danz, Frank Hüber, Dorothea Kübler, Lydia Mechtenberg, Julia Schmid SP II 2013-203
'I'll do it by myself as I knew it all along': On the failure of hindsight-biased principals to delegate optimally

David Hugh-Jones, Morimitsu Kurino, Christoph Vanberg SP II 2013-204
An Experimental Study on the Incentives of the Probabilistic Serial Mechanism

Yan Chen, Onur Kesten SP II 2013-205
From Boston to Chinese Parallel to Deferred Acceptance: Theory and Experiments on a Family of School Choice Mechanisms

Thomas de Haan, Roel van Veldhuizen SP II 2013-206
Willpower Depletion and Framing Effects

Christine Binzel, Dietmar Fehr SP II 2013-207
Giving and sorting among friends:
evidence from a lab-in-the-field experiment

Volker Benndorf, Dorothea Kübler, Hans-Theo Normann SP II 2013-208
Privacy Concerns, Voluntary Disclosure of Information,
and Unravelling: An Experiment

Rebecca B. Morton, Marco Piovesan, Jean-Robert Tyran SP II 2013-209
Biased Voters, Social Information, and Information
Aggregation Through Majority Voting

Roel van Veldhuizen SP II 2013-210
The influence of wages on public officials' corruptibility: a laboratory investigation

Jana Friedrichsen SP II 2013-211
Image concerns and the provision of quality

Research Unit: **Economics of Change**

Luisa Herbst, Kai A. Konrad, Florian Morath SP II 2013-301
Endogenous Group Formation in Experimental Contests

All discussion papers are downloadable:
http://www.wzb.eu/en/publications/discussion-papers/markets-and-choice

Kai A. Konrad, Florian Morath SP II 2013-302
Evolutionary Determinants of War

Armin Falk, Nora Szech SP II 2013-303
Organizations, Diffused Pivotality
and Immoral Outcomes

Maja Adena, Steffen Huck, Imran Rasul SP II 2013-304
Charitable Giving and Nonbinding Contribution-Level Suggestions.
Evidence from a Field Experiment

Dominik Rothenhäusler, Nikolaus Schweizer, Nora Szech SP II 2013-305
Institutions, Shared Guilt, and
Moral Transgression

Dietmar Fehr, Steffen Huck SP II 2013-306
Who knows it is a game? On rule understanding, strategic awareness
and cognitive ability

Maja Adena, Michal Myck SP II 2013-307
Poverty and Transitions in Health

Friedel Bolle, Jano Costard SP II 2013-308
Who Shows Solidarity with the Irresponsible?

Kai A. Konrad SP II 2013-309
Affection, Speed Dating and Heart Breaking

Maja Adena, Ruben Enikolopov, Maria Petrova, SP II 2013-310
Veronica Santarosa, and Ekaterina Zhuravskaya
Radio and the Rise of the Nazis in Pre-war Germany

Perdro Dal Bó, Guillaume R. Fréchette SP II 2013-311
Strategy Choice In The Infinitely Repeated Prisoners' Dilemma

Steffen Huck, Gabriele K. Lünser, Jean-Robert Tyran SP II 2013-312
Price Competition and Reputation in Markets
for Experience Goods: An Experimental Study

Rune Midjord, Tomás Rodríguez Barraquer, Justin Valasek SP II 2013-313
Over-Caution of Large Committees of Experts

Steffen Huck, Julia Schmid, Brian Wallace SP II 2013-314
Price Framing

Pio Baake, Steffen Huck SP II 2013-315
Crop Failures and Export Tariffs

WZB Junior Research Group: **Risk and Development**

Lubomír Cingl, Peter Martinsson, Hrvoje Stojic, Ferdinand M. Vieider SP II 2013-401
Separating attitudes towards money from attitudes towards probabilities:
stake effects and ambiguity as a test for prospect theory

All discussion papers are downloadable:
http://www.wzb.eu/en/publications/discussion-papers/markets-and-choice