Third-Party Surveillance and Third-Party Punishment in Trust Games:

Experimental Evidence on Prevailing Trust Norms and Their Enforcement

Lisa Einhaus
einhau01@uni-passau.de

Date: January 31, 2012

University of Passau

Winter Term 2011/12

Seminar in Experimental Economics

Content

List of Tables

List of Figures

1 Introduction

The concept of trust has been investigated in the context of different disciplines. Sometimes, it is found to be a *magic bullet* enhancing the performance of economies, firms, institutions and even of civic groups. Whereas cooperation between friends or business partners may rather be explained by the repetition of transactions and reputation building than by efforts of trust, unraveling motives behind *pure* trusting behavior seems less trivial (La Porta et at. 1997, 333f). Pure trust is required in anonymous one-shot exchanges, not only bearing a possibility of benefit, but also of loss in case others defect (Cox 2004, 263). Why is it that some populations exhibit this kind of trust, while others don't?[1] Might trusting behavior be driven by common norms of a society, or put differently, do people exhibit trust owing to a certain normative obligation they feel?

In experimental economics social norms have been modeled by means of third-party[2] punishment. As opposed to sanctions by a harmed second-party, who might be influenced by retaliation or affection, sanctions imposed by a third-party are seen to qualify as a yardstick whereby one can judge whether a certain behavior may be considered a normative standard (Ohtsubo et al. 2010, 259). The paper at hand employs third-party punishment in the context of a trust game in order to elicit prevailing norms with respect to trusting behavior. However, since experimental studies sometimes find that monetary sanctions crowd out intrinsic motivation and have a negative impact on cooperative behavior (Falk and Kosfeld 2004, Camerer 2003, 96f) an alternative enforcement mechanism, namely a surveilling third-party, who ex post observes the performance of co-players, is analyzed, too.

[1] For a brief inter-ethnic survey see Camerer 2003, 86.
[2] A third-party is here defined as an outside subject, whose monetary payoff is not directly affected by co-player performance.

2 Underlying Economic Theory and Previous Research

2.1 The Trust Game

In economics trust is frequently measured in a one-shot, two-subject trust game, sometimes also called investment game. The game was introduced by Berg, Dickhaut and McCabe (1995) and involves a first mover, the trustor, and a second mover, the trustee. Both receive an initial endowment of equal stake size. The trustor may pass any amount of his endowment to the trustee, with investment being multiplied before transferred. The trustee then decides how much of his resources to pass back to the trustor. The only Nash equilibrium is zero investment: since by assumptions of game theory subjects maximize own payoffs, the first mover should not expect any transfer by the trustee at all and in anticipation chooses not to invest in the first place. If, nevertheless, a positive sending by the first mover occurs, the amount transferred is often seen as a proxy for measuring trust (Camerer 2003, 85)[3].

Berg et al. (1995) find that trustors on average invest half of their endowment. Replicable behavior is reported by several studies.[4] Can prevailing normative standards account for this kind behavior? This assumption is backed by a social history treatment conducted by Berg et al. (1995), finding that knowing how others performed during earlier games - thus fostering expectations - reinforces inclinations towards trusting behavior.

2.2 Third-Parties and Social Norms

When it comes to third-parties, researchers previously focused on the enforcement of fairness or egalitarian distribution norms. In dictator games, prisoner's dilemmas and public goods games it was found that the strength of punishment is not random, but somewhat proportional to the norm deviation (Fehr and Gächter 2000, Fehr and Fischbacher 2004, Ottone 2005). A similar correlation can be found for the violation of reciprocity and honesty norms in trust games (Ohtsubo et al. 2010, Charness et al. 2008). With respect to the enforcement of reciprocity norms, Charness et al. (2008) find that transfers by trustees increase and trustors, too, adjust their strategy by investing more when a third-party may punish the trustor for possible norm violation. The question may be raised, if the enforcement of a trust norm, directly addressing first movers with a threat of punishment, influences trusting behavior.

[3] Researchers have found that besides trust other motivations might be relevant for positive transfers of trustors. With regard to the findings of this experiment, they are discussed in section 7.
[4] For a brief survey see Camerer 2003, 84-86.

Bicchieri et al. (2011) investigate prevailing trust norms by means of a survey. Providing the description of a trust game, it is asked whether people believe that most other people punish a lack of trust. Since answers neither for transfers between friends, nor between strangers are affirmative, the authors conclude that trusting may not be considered a norm. However, the study does not involve experimental game play and focuses on the willingness of third-parties to sanction first movers who fail to trust. This study, on the contrary, also takes into account changes of first mover behavior when a third-party is introduced.

A second novelty of this paper is examining the effects of a bystanding third-party, neither entitled to impose monetary, nor informal sanctions[5]. Evidence from a field study by Bateson et al. (2006) suggests that subtle cues of observation (i.e. pictures of human eyes) increase public good contributions. This may sense that even though - as opposed to formal and informal punishment - feedback by the third-party is not possible, subjects might subconsciously be focused on common norms. As compared to formal and informal sanctions, a bystanding third-party holds the characteristic that it is self-enforcing.

3 Experimental Design

The experiment's set up is similar to the Berg et al. (1995) investment game described in section 2.1. The initial endowment is 10 Passau Taler (PT), investment choices are 0, 2, 4, 6, 8 or 10 PT and investment is tripled before passed to the trustee. The following three treatments are played as a within-subject design (for a schematic overview see Figure 1a):

Treatment 1 (T1) resembles a classical two-player investment game.

Treatment 2 (T2) introduces a passive third-party, the observer. He is inactive, but learns about investment and payback decisions of co-players at the end of the experiment. There is full information, i.e. all probands know about the presence of the observer before T2 begins. The observer receives a show up fee equal to the initial endowment of other players.

Treatment 3 (T3) entitles the third-party, the punisher, to ex post valuate trustor performance by reducing his payoff by 0, 1, 2, 3 or 4 PT.[6] As in T2, all players are informed about this modification. Further, after respective decisions have been submitted, a *belief stage* is added for both the trustor and the punisher. The trustor is asked about the payoff reduction he

[5] Examples of informal sanctions are the expression of opinion or the distribution of disapproval points. For the cooperation enhancing power of disapproval points see Masclet et al. (2003).
[6] Note that the reduction is not transferred to the third-party, i.e. does not yield monetary payoff.

expects for his investment. This allows finding out about the trustor's perception of prevailing norms. Similarly, the punisher is asked about his belief regarding the average valuation of fellow third-parties in order to test, if the norm he enforces is valid with respect to the observed population. For revealing true beliefs, the punisher's payoff is dependent on the accurateness of his beliefs, i.e. the deviation of his expectation from the average valuation chosen by other punishers (10-|$\bar{o}$-p|, with 10 PT being his initial endowment, $\bar{o}$ the average sanction imposed by other third-parties and p his belief regarding average punishment for a given investment level). Trustor payoff is not adjusted, assuming that he already formed his belief earlier when deciding upon his investment level under consideration of possible sanction.

Answers of the punisher are collected using the strategy method, with the investment level actually chosen by the trustor being payoff-relevant. Employing the strategy method is essential in order to avoid inter-temporal carryover effects that can arise between treatments due to the within-subject design. For the same reason investment, repayment and payoff (and for T3 punishment) choices are not displayed until the end of the experiment and subjects are not informed beforehand about the number of treatments played.

As this paper focuses on the performance of first movers and third-parties, a matching protocol is employed. In each session only one subject acts as a trustee and is matched with multiple trustors and an equal number of third-parties. Therefore, trustee decisions are collected via the strategy method and his payoff is calculated and displayed as an average. Roles are assigned randomly and - similar to the matching protocol - kept constant during all treatments. Due to the fact that the belief stage requires at least two trustor-punisher pairs playing simultaneously, the minimum number of players per session is five, with a higher odd number of subjects being possible, too. Neutral wording is chosen to rule out framing effects.

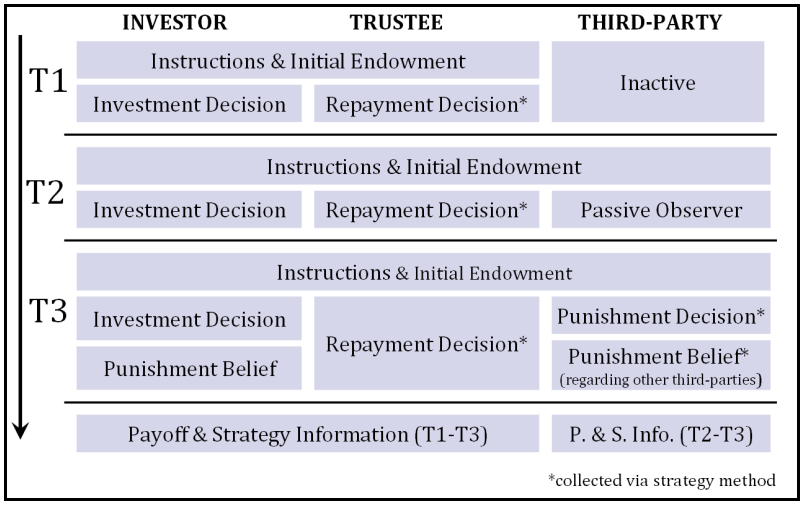


Figure 1: Schematic Depiction of Experimental Design

4 Hypotheses

Based on the preceding findings, the following hypotheses shall be tested:

H 1: In the presence of a third party, trustors change their contributions in accordance with a trust norm, i.e. amounts sent by first movers are higher in T2 and T3 when compared to T1.

H 2: The effect of a punisher (T3) is stronger than of an observer who functions as a bystander (T2), since formal sanctions are often considered essential for the enforcement of social norms (Bendor and Swistak 2001, 1494).

H 3: Trust is a norm that is enforced rather by the belief of sanction than its factual imposition, i.e. trustors expect higher punishment than actually imposed. This may be due to risk aversion.

Under the assumption that third-parties effectively enforce prevailing social norms, overestimation of factual sanction might extend to the beliefs of third-parties, too.

5 Procedures

The experiment was designed in the framework of the seminar *Experimental Economics* at the University of Passau. It was programmed using the software *z-Tree* (Fischbacher 2007). On November 21st, 2011 five sessions (10am, 12am, 2pm, 4pm, 6pm), with 9-15 participants each and 57 subjects participating in total, took place in one of the University's computer labs (WiWi 030). The five sessions were organized jointly with two fellow students. A brief welcoming speech was held at the beginning of each session, while experiment-specific information was displayed only on screen as part of the instructions. Each experiment lasted about 15-20 minutes, with each session taking 45-60 minutes. This experiment was conducted first in all five sessions. Participation was voluntary and subjects were recruited on campus, via flyers and announcements, as well as online, via facebook.com and the University online platform *stud.ip,* which was also used for in advance registration.

6 Analysis of Experimental Results

Of the 57 subjects who participated in the experiment, 33 (57.9%) are female and 24 (42.1%) male. 80% of the participants are enrolled students, most of them study at the faculties for humanities (43.9%) and economics (42.1%), followed by the faculties for law (8.8%) and for computer science/mathematics (5.3%). By the matching protocol described in section 3, each session consisted of one trustee and multiple trustor-third-party pairs. In total, five subjects took the role of a trustee and each 26 the role of a trustor and a third-party.

6.1 Hypothesis 1

The introduction of a third-party has induced a change in trustor behavior, though at first glance no clear pattern in favor of trust norm conform behavior can be identified: sending of 0 and 10 PT has increased from T1 to T2 and T2 to T3, investment of 4 and 8 PT decreases continuously between T1 and T3, while transfers of 2 and 6 PT increase when an observer is present, but slightly drop back again when punishment is feasible (Table 1). Considering average investment (Figure 2), H1 can be confirmed: sending in the baseline treatment T1 is lower (6.3 PT) than in T2 (6.7 PT) and T3 (6.9 PT).

		Treatment 1	Treatment 2	Treatment 3
Investment Level	0	3.8 % (1)*	7.7% (2)	11.5% (3)
	2	11.5% (3)	3.8% (1)	7.7% (2)
	4	26.9% (7)	15.4% (4)	7.7% (2)
	6	11.5% (3)	26.9% (7)	19.2% (5)
	8	15.4% (4)	11.5% (3)	7.7% (2)
	10	30.8% (8)	34.6% (9)	46.2% (12)

Table 1: Trustor Choices by Treatment (total) *absolute frequencies noted in parentheses

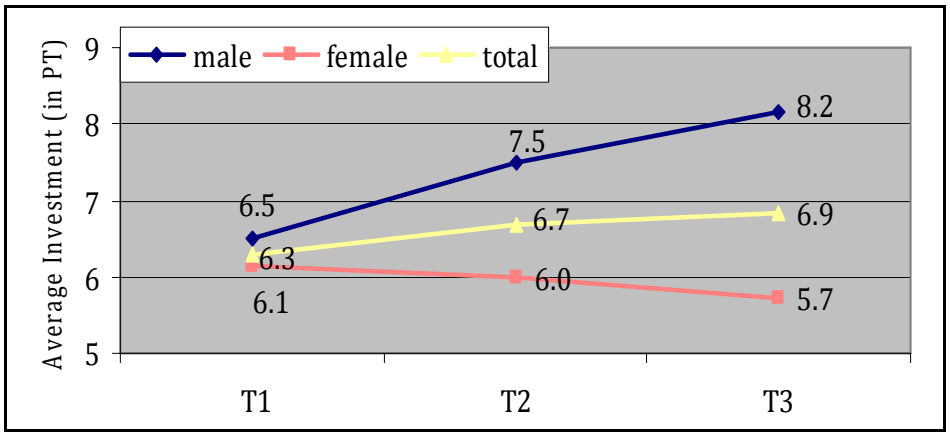


Figure 2: Average Investment in Treatments T1, T2 and T3 (total and by gender)

When looking at subject performance in account of gender, this finding has to be modified. Male participants undertake high investments more often when a third-party is present (Table 2) and on average invest 6.5 PT in T1, 7.5 PT in T2 and 8.2 PT in T3 (Figure 2). A diametrical pattern is observed for female participants, who on average invest 6.1 PT in T1, 6.0 PT in T2 and 5.7 PT in T3 (Figure 2). This does not refute the relevance of third-parties in itself, but rather suggests that there is no collective orientation regarding trust: men - as opposed to women - might simply perceive trust to be a prevailing norm and anticipate accordingly.

		Treatment 1		Treatment 2		Treatment 3	
		male	female	male	female	male	Female
Investment Level	0	8.3% (1)*	0.0% (0)	8.3% (1)	7.1% (1)	8.3% (1)	14.3% (2)
	2	8.3% (1)	14.3% (2)	0.0% (0)	7.1% (1)	0.0% (0)	14.3% (2)
	4	16.7% (2)	35.7% (5)	8.3% (1)	21.4% (3)	0.0% (0)	14.3% (2)
	6	8.3% (1)	14.3% (2)	16.7% (2)	35.7% (5)	16.7% (2)	21.4% (3)
	8	33.3% (4)	0.0% (0)	25.0% (3)	0.0% (0)	16.7% (2)	0.0% (0)
	10	25.0% (3)	35.7% (5)	41.7% (4)	28.6% (4)	58.3% (7)	35.7% (5)

Table 2: Trustor Choices by Treatment (by gender) *absolute frequencies noted in parentheses

This argument is supported when taking a closer look at gender-specific punishment decisions of third-parties. As laid out in section 2.2, previous studies found that punishment is fairly proportional to norm violation. Figure 3 illustrates that men punish low investment stronger and high investment weaker than women. On average women sanction sending of 6, 8 and 10 PT with 2.0 to 2.1 PT strongest, while sanction for 2 PT- and 0 PT-investment is relatively low (1.5-1.6 PT). Punishment by male third-parties indicates a different pattern: while investment of 0 and 2 PT renders in punishment of 3.5 PT, sanctioning falls with higher levels of investment to 0.5 PT for a 10 PT-transfer.

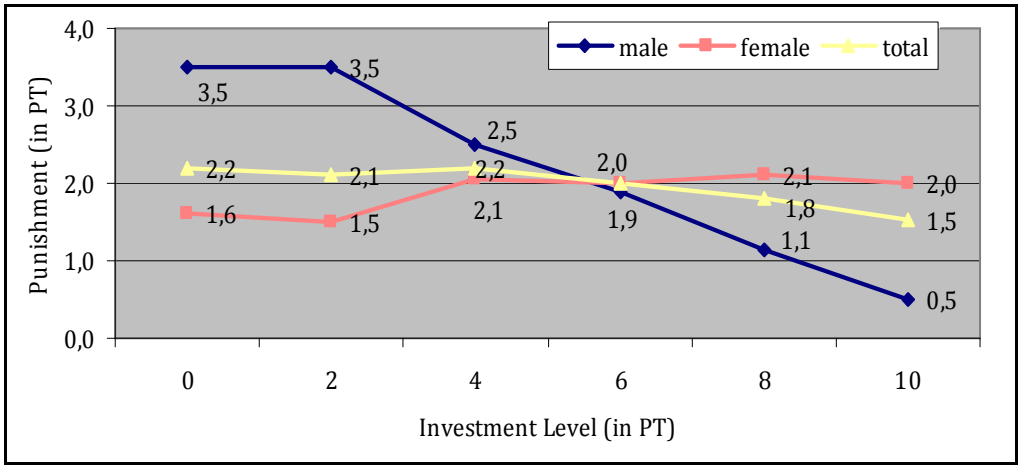


Figure 3: Average Third-Party Punishment (total and by gender)

A 0 PT-transfer is not punished at all by half of female third-parties, while 88% of male third-parties choose 4 PT, the highest possible sanction. Considering a 10 PT-investment, 88% of male third-parties do not punish at all, compared to only 39% of female punishers (Figure 4). These findings further indicate that women have less of a normative concern for trusting behavior, while men view a lack of trust as a norm violation.

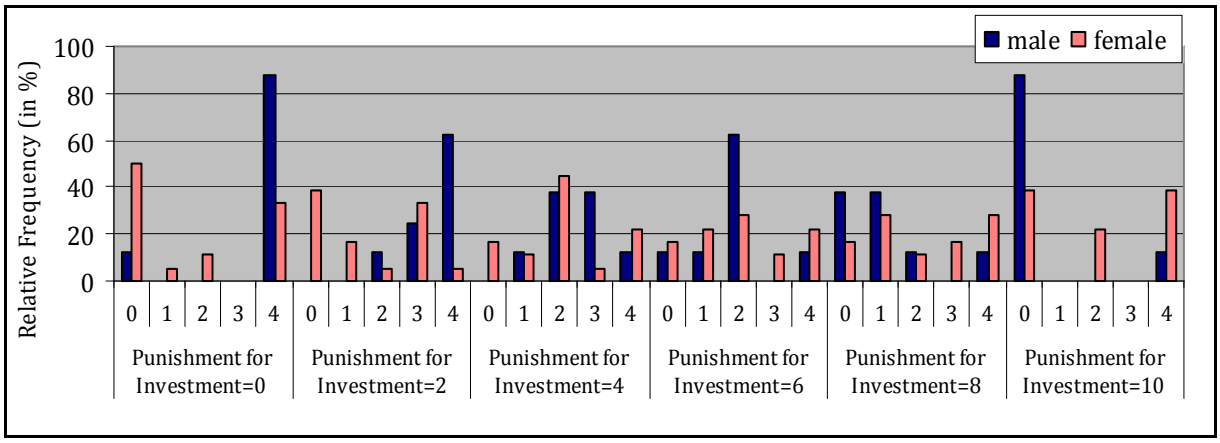


Figure 4: Relative Frequency of Punishment for Given Investment Levels (by gender)

6.2 Hypothesis 2

Recalling the findings on overall, male and female average investment (section 6.1), H 2 can be confirmed: the effect of a punishing third-party on trustor decisions is stronger than of a passive third-party (Figure 2). It is quite striking though that a passive observer enhances contributions of first movers fairly strong.[7]

To test the significance of this finding, pairwise Wilcoxon signed-rank tests (T1-T2, T1-T3, T2-T3) were calculated using SPSS. When considering aggregated data, the investment level of trustors was not significantly influenced by the introduction of a third-party.[8] When examining genders separately, it is found that for male trustors third-party presence/ punishment can significantly account for the observed change in trusting behavior (p-value= 0.034 for T1-T2 and 0.034 for T1-T3). However, "activating" the third-party does not bear a significant additional effect on trustor behavior (p-value=0.68 for T2-T3).

[7] Note that from a game theoretic perspective an observer, who does not affect players' monetary payoff, should not influence their strategy.
[8] (for all pairwise tests p-value>0,05)

6.3 Hypothesis 3

When comparing average expected punishment of trustors and third-parties with factual imposition (Figure 5), it cannot be confirmed that prevailing norms are enforced rather by the belief of sanctions than their factual imposition. Neither trustors, nor third-parties consistently overestimate the scope of punishment for all investment levels.

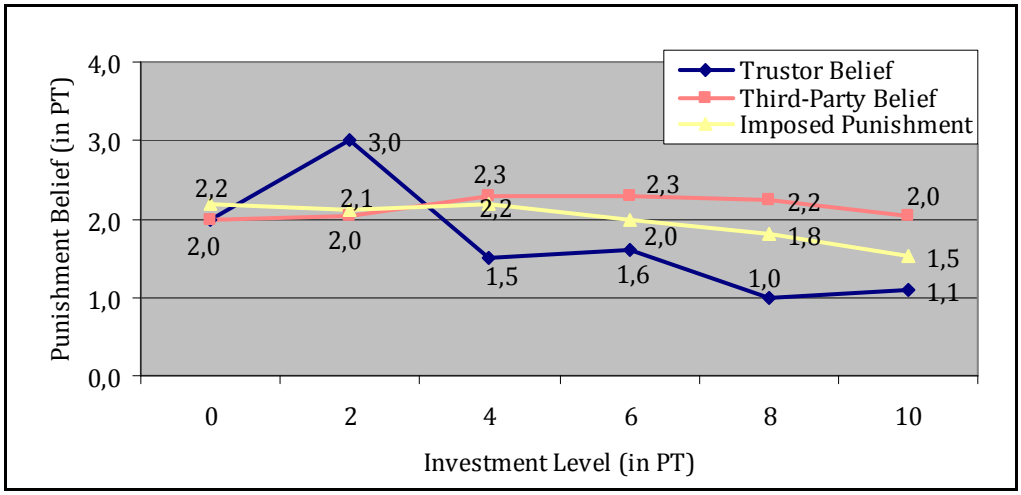


Figure 5: Comparison of Beliefs and Imposed Punishment

However, Figures 6 and 7 illustrate that the observed gender bias (section 6.1) persists in beliefs, too: while female trustors expect an average sanction of only 1 PT for a 0 PT-investment, male trustors expect 4 PT (actual mean sanction is 2.2 PT).[9] For a 10 PT-investment this relationship is reverse: male expect only 0.7 PT-punishment, while women belief that a sanction of 1.6 PT will occur (actual mean punishment is 1.5 PT) (Figure 6). Similar, but less strong deviations of beliefs from imposed sanctions can be observed for third-parties (Figure 7).

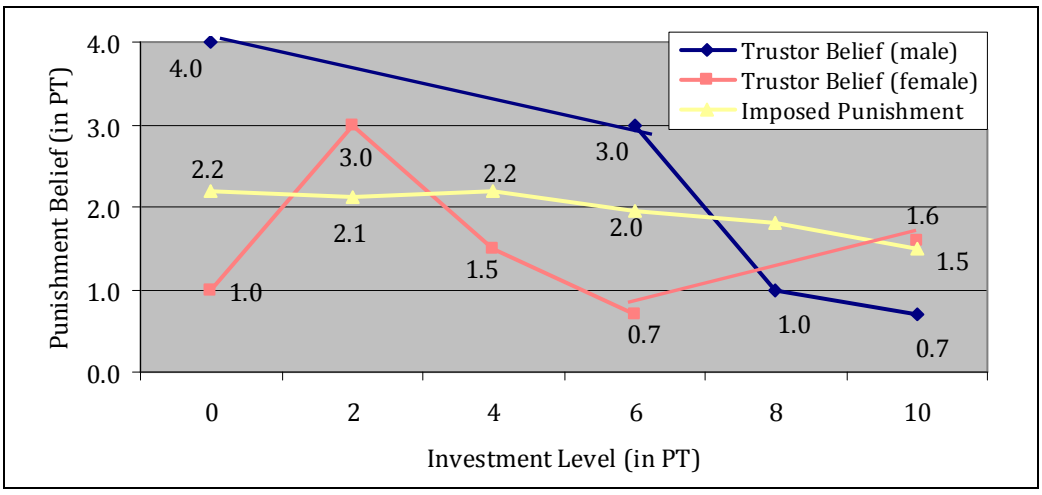


Figure 6: Trustor Beliefs Regarding Average Punishment (total and by gender)

[9] Regarding the calculation of averages, note that only 8 third-parties are male, while 18 are female. Game play was only anonymous with regard to the performance of the participants. Subjects were able to see who else participated in the ongoing session, with gender distribution being fairly stable throughout sessions.

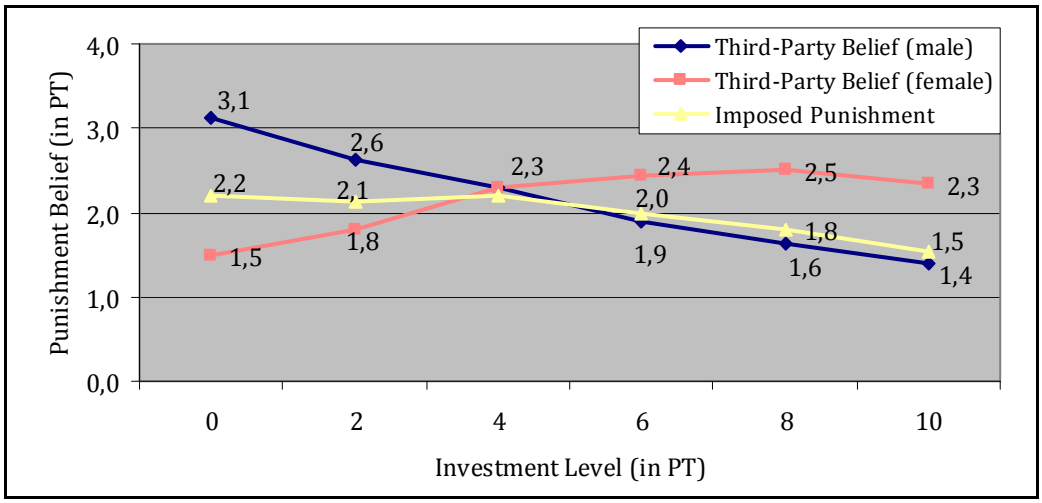


Figure 7: Third-Party Beliefs Regarding Average Punishment (total and by gender)

One reason why a gender bias with respect to attitudes towards trust persists in beliefs might be that both genders fail to recognize the gender-specific concerns for trusting behavior and in consequence fail to form good beliefs with respect to the entire observed population. Figures 8 and 9 compare trustor and punisher beliefs with factual punishment by third-parties of the same gender and signal evidence in favor of this train of thought. With the exception of female trustors, beliefs of all other groups are relatively accurate.

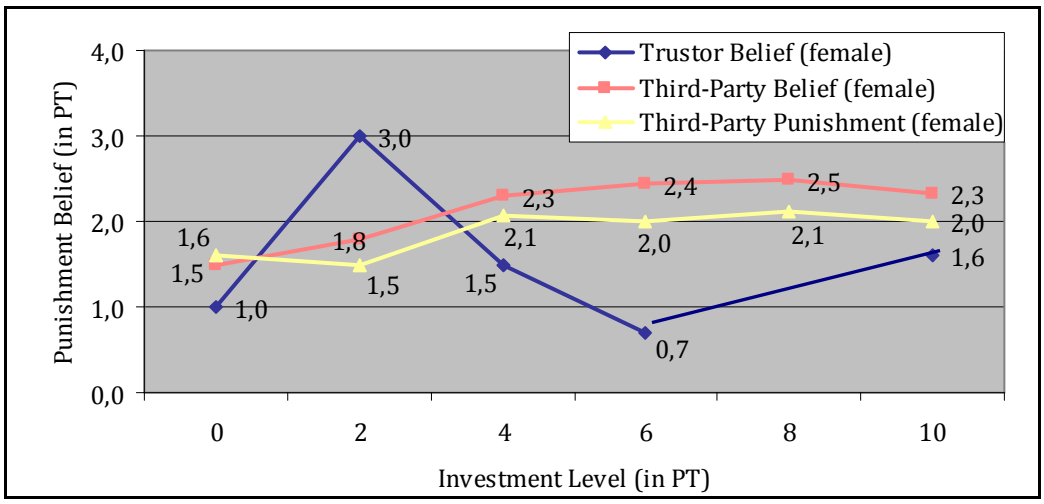


Figure 8: Third-Party Punishment, Trustor Beliefs and Third-Party Beliefs (female)

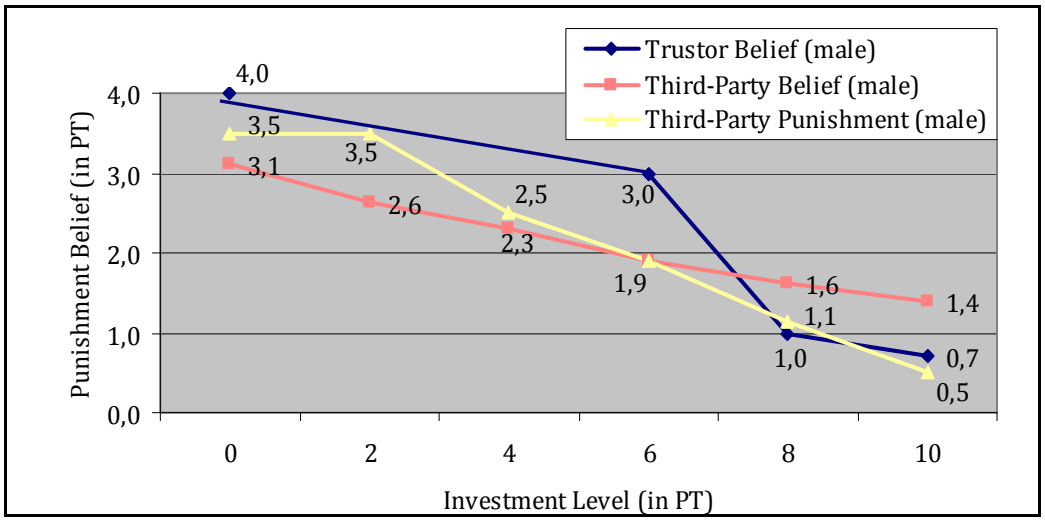


Figure 9: Third-Party Punishment, Trustor Beliefs and Third-Party Beliefs (male)

6.4 Trustee Behavior

Though not being the focus of the experiment, trustee behavior was recorded as followed: paybacks increase from 5.5 PT in T1 to 6.5 PT in T2 and 6.9 PT in T3. Wilcoxon signed-rank tests (T1-T2, T1-T3, T2-T3) were calculated, showing that for none of the six investment levels the change in repayments can be explained through the introduction of a third-party.[10]

7 Discussion of Experimental Results

Behavioral scientists have found that the concept of trust is closely interrelated with reciprocity. As Cox (2004) puts it: "Trust is inherently a matter of the beliefs that one agent has about the behavior of another" (Cox 2004, 263). Also, subjects may not only have self-regarding, but also other-regarding preferences, suggesting that positive transfers may be induced by altruistic giving (Cox 2002, 334). Decomposing trust and altruism by comparing first mover behavior in dictator and investment games, Cox (2002) observes no significant trusting behavior, while Cox (2004) comes to the conclusion that significant amounts of investment stem from trust. Thus, the observed diametrical behavior of males and females may not solely be explained by different normative standards towards trust, but possibly also towards reciprocity and other-regarding preferences.

Indeed, previous experimental studies suggest that men do exhibit relatively strong concerns for reciprocity, while actions by females are rather driven by other-regarding preferences

[10] (p-value >0,05 for all investment levels)

(Cox 2002, Lambsdorff and Frank 2011, Buchnan et al. 2002, Chaudhuri and Gangadharan 2003). Applying these findings to the study at hand, this might sense that when there is norm-enforcement by a third-party, males do not only perceive the enforcement of a trust norm, but also - or maybe exclusively - of a reciprocity norm: since the magnitude of possible reciprocity stems from initial trustor transfers, men might expect punishment of low investment (restricting reciprocity) and in anticipation rationally increase their initial transfer when a third-party is present. As seen in section 6.1, sanctioning by male third-parties is in line with this train of thought. Turning to females, if first mover transfers do dependent on other-regarding preferences, their intrinsic motivation might be at conflict when extrinsic enforcement mechanisms are present. This would account for the observed decreasing contributions of women between treatments.

Falk and Koesfeld (2004) find that an agent's performance is negatively influenced, if his choice set is limited by a principal. Female trustors might similarly perceive both a surveilling and a punishing third-party as limiting their choice to trust or not to trust, in consequence lowering contributions in T2 and T3. Alternatively, women might perceive "spending trust" as a cognitive or emotionally challenging task, showing signs of fatigue as treatments proceed, as sometimes observed independent of gender in the context of charitable contributions (Camerer 2003, 60).

8 Limitations

Since most of the participants were students at the University of Passau, findings might be limited to the analyzed subject pool. Further, participation was voluntarily, whereby a certain self-selection might have occurred. The fraction of students with game theory knowledge was disproportionately high and payoffs were hypothetical, possibly having influenced results.

9 Conclusion

The paper at hand set off with the question, if people exhibit trust due to prevailing norms and social obligations. The analysis of the experimental data has revealed three striking results:

First, men and women do not share common, but rather gender specific norms with regard to trusting behavior. Under norm-enforcement males exhibit more trust than women, though this causality could - instead of being driven by a trust norm - also stem from differing concerns for reciprocity and other-regarding preferences (see section 7). Decomposing the driving forces behind trusting behavior remains a challenge for future research.

Second, a gender bias in attitudes towards trust holds true for beliefs, too. Trustor and third-party beliefs are not accurate when compared to factual punishment imposed on the whole, but are rather in line with actions taken by the same gender. This suggests that neither males nor females recognize the existing gender bias towards trust.

Third, when compared to mere surveillance, entitling a third-party to impose monetary sanctions only has an insignificant extra effect for the power of norm enforcement. This signals that when it comes to social norms and their enforcement, subjects might not only be concerned with possible punishment, but also with behaving conform to expectations of others. It would be interesting to examine whether this powerful role of passive third-parties extends to different contexts and different norms.

References

Bateson, M., Nettle, D. and Roberts, G. (2006). Cues of being watched enhance cooperation in real-world setting. Biology Letters 2, 412–414.

Bendor, J. and Swistak, P. (2001). The evolution of norms. American Journal of Sociology 106, 1493-1545.

Berg J., Dickhaut J. and McCabe K. (1995). Trust, Reciprocity, and Social History. Games and Economic Behavior 10, 122–42.

Bicchieri, C., Xiao, E. and Muldoon, R. (2011). Trustworthiness is a social norm, but trusting is not. Politics, Philosophy & Economics 10(2), 170-187.

Buchnan, N., Croson, R. and Solnick, S. (2008). Trust and gender: an examination of behavior and beliefs in the Investment Game. Journal of Economic Behavior & Organization, 68, 466-476.

Camerer, C. (2003). Behavioral Game Theory: Experiments in Strategic Interaction.

Charness, G., Cobo-Reyes, R. and Jimenéz, N. (2008). An investment game with third-party intervention. Journal of Economic Behavior & Organization 68, 18-28.

Chaudhuri, A., Gangadharan, L. (2003). Gender Differences in Trust and Reciprocity. The University of Auckland, Economics Working Papers.

Cox, J. (2002). Trust, reciprocity, and other-regarding preferences: groups vs. individuals and males vs. females. In Zwick, R. and Rapoport, A. (Eds.), Experimental Business Research, 331–349.

Cox, J. (2004). How to identify trust and reciprocity. Games and Economic Behavior 46, 260-281.

Falk, A. and Kosfeld, M. (2004). Distrust - The Hidden Cost of Control. IZA Discussion Paper 1203.

Fehr, E. and Gächter, S. (2000). Cooperation and Punishment in Public Goods Experiments. American Economic Review 90(4), 980-994.

Fehr, E. and Fischbacher, U. (2004). Third-party punishment and social norms. Evolution and Human Behavior 25, 63–87.

Fischbacher (2007). z-Tree: Zurich Toolbox for Ready-made Economic Experiments. Experimental Economics 10(2), 171-178.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A. and Vishny, R (1997). Trust in large organizations. American Economic Association Papers and Proceedings 87(2), 333–338.

Lambsdorff, J. Graf and Frank, B. (2011). Corrupt reciprocity – Experimental evidence on a men's game. International Review of Law and Economics 31, 116-125.

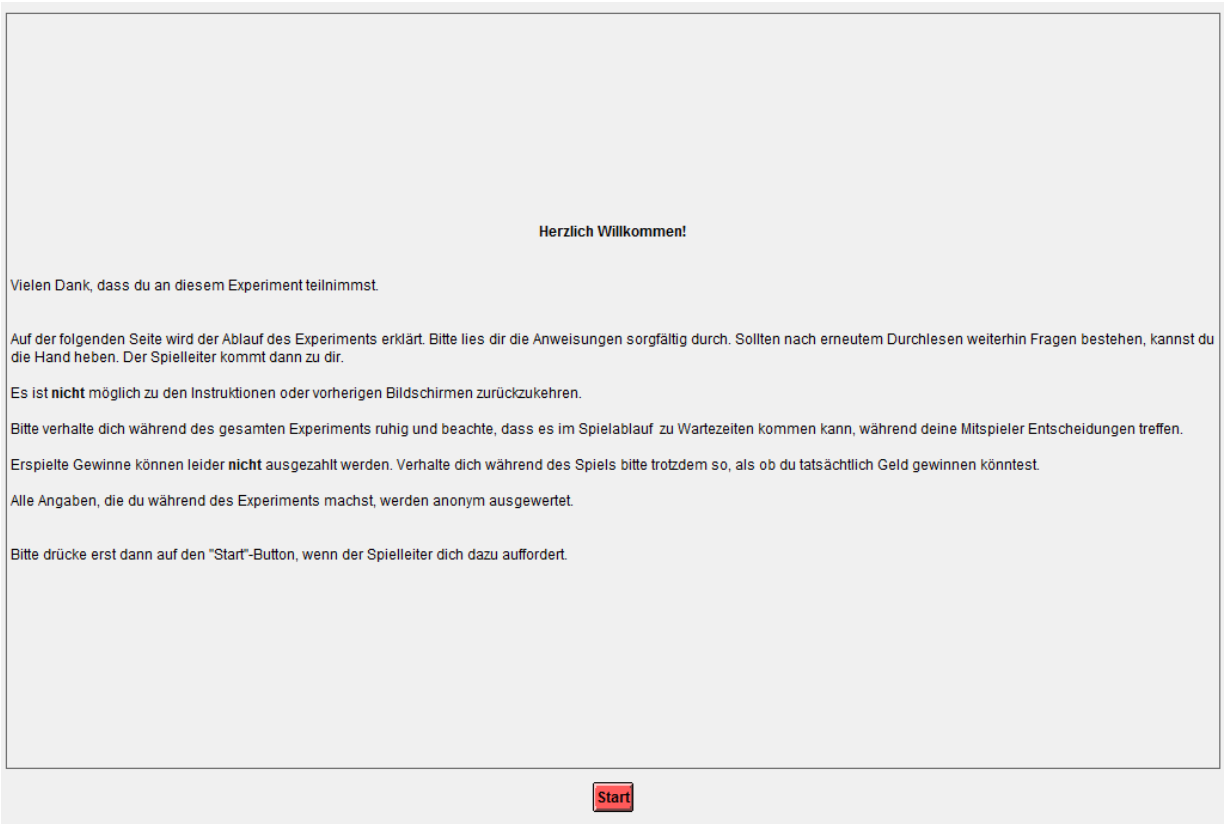
Masclet, D., Noussair, C., Tucker, S. and Villeval, M.-C. (2003). Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism. American Economic Review 93(1), 366-380.

Ohtsubo, Y., Masuda, F., Watanabe, E. and Masuchi, A. (2010). Dishonesty invites third-party punishment. Evolution and Human Behavior 31, 259-264.

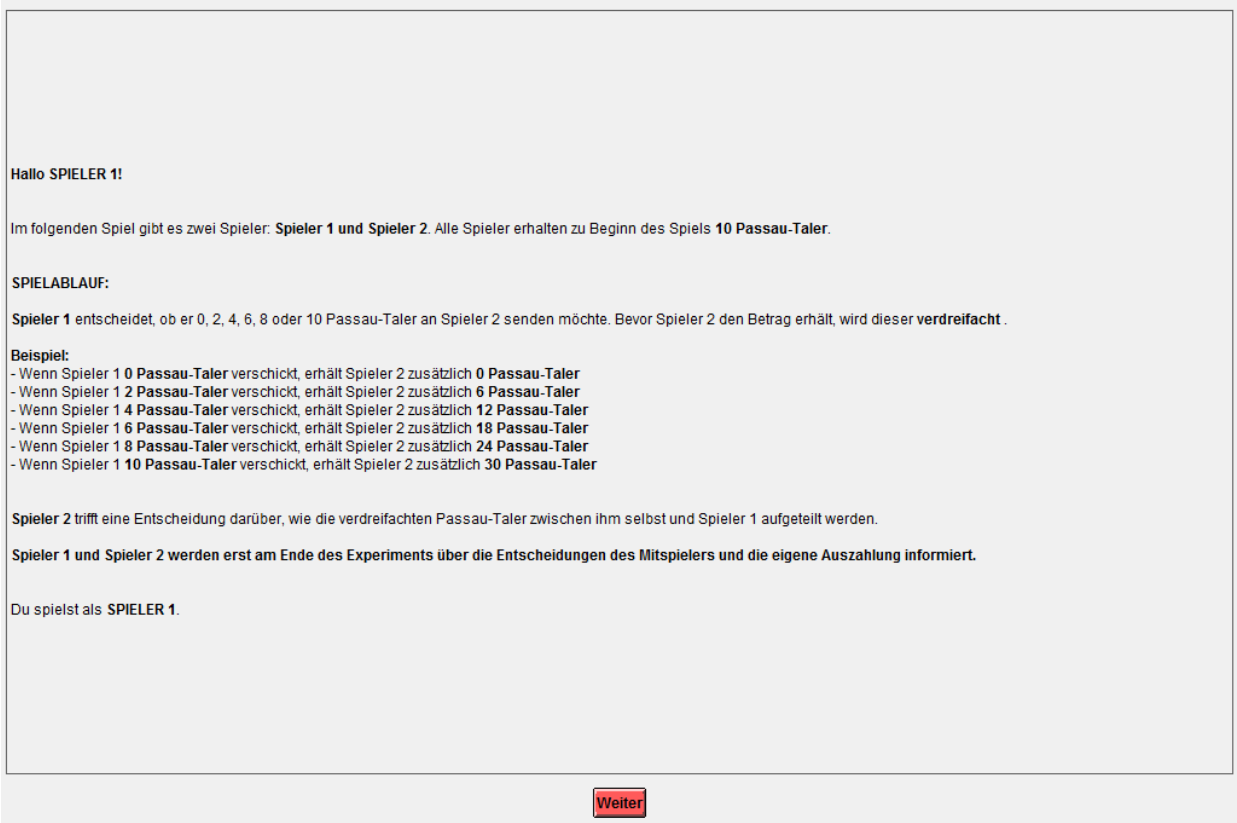
Ottone, S. (2005). Transfers and altruistic punishments in Solomon's game experiments. Department of Public Policy and Public Choice - POLIS, Working Paper No.57.

Appendix

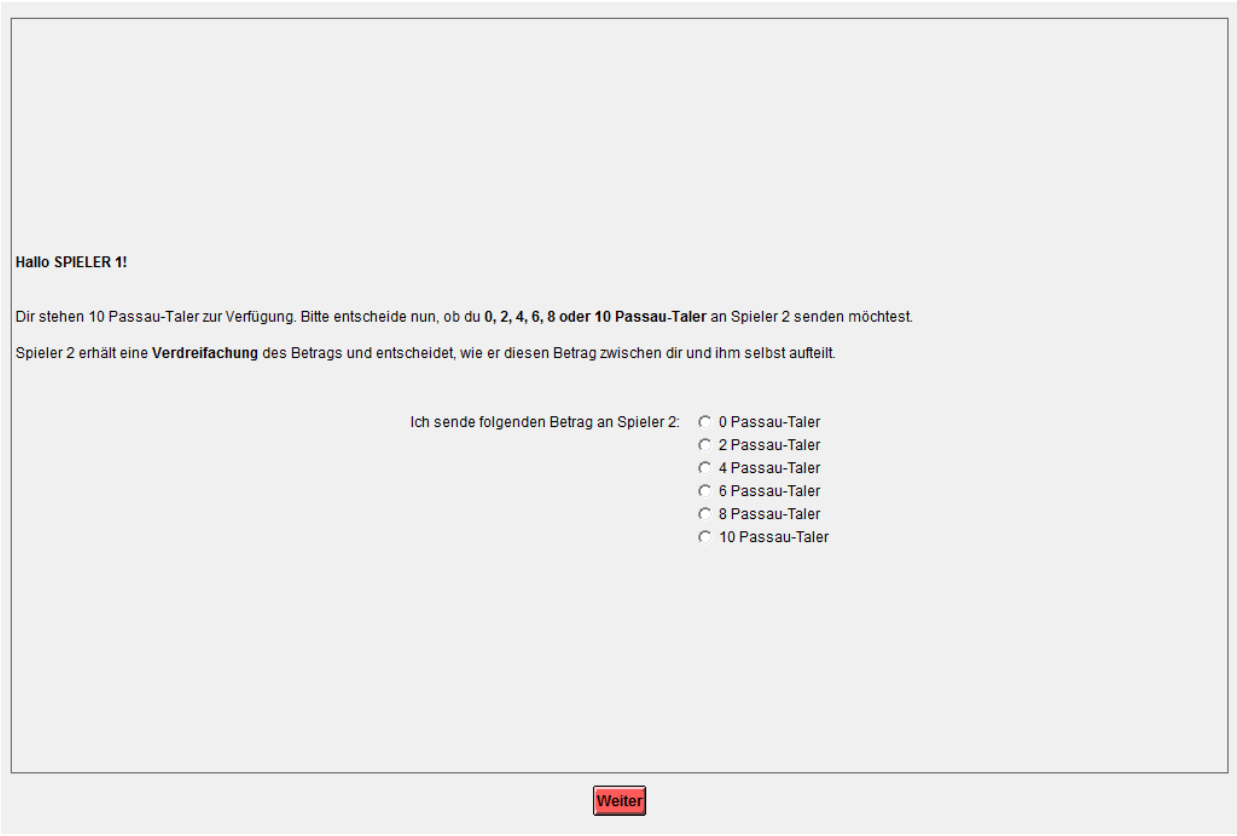
Screenshots



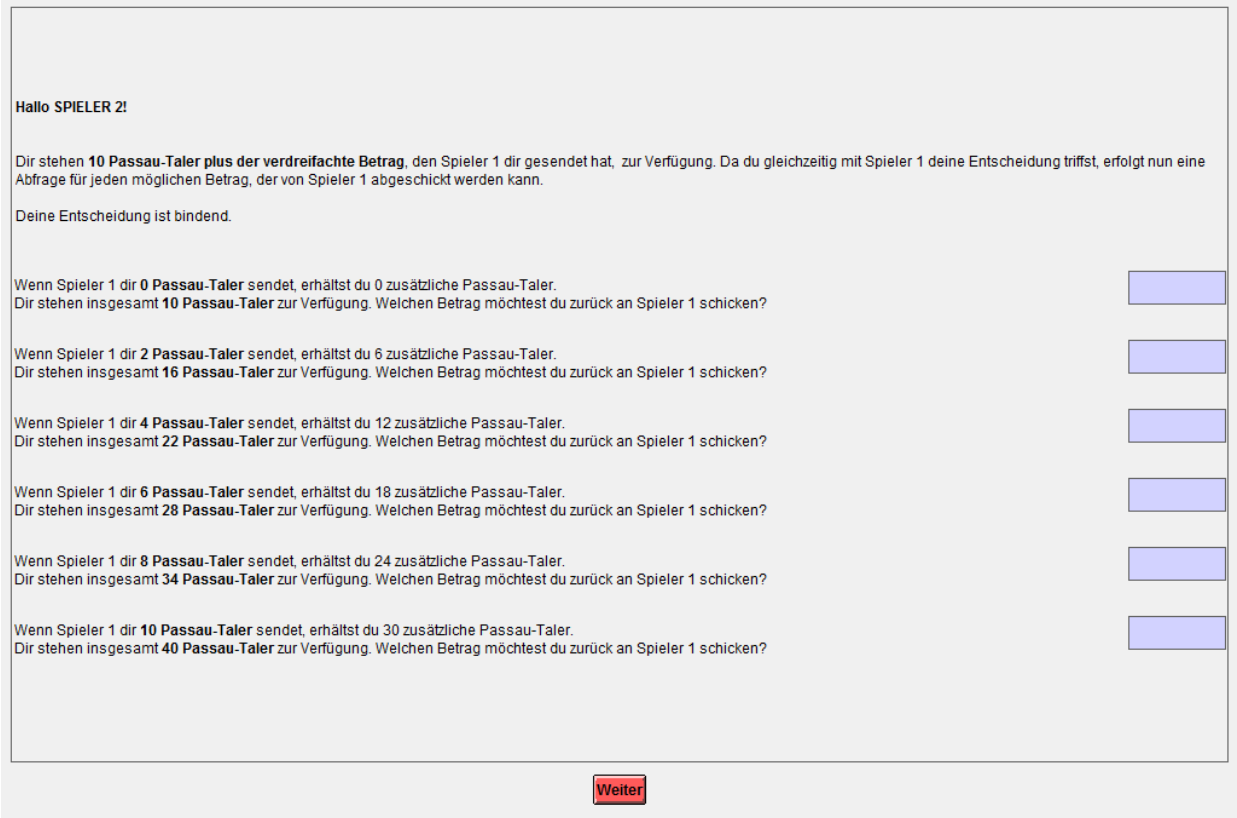
Screenshot 1: Welcome Screen



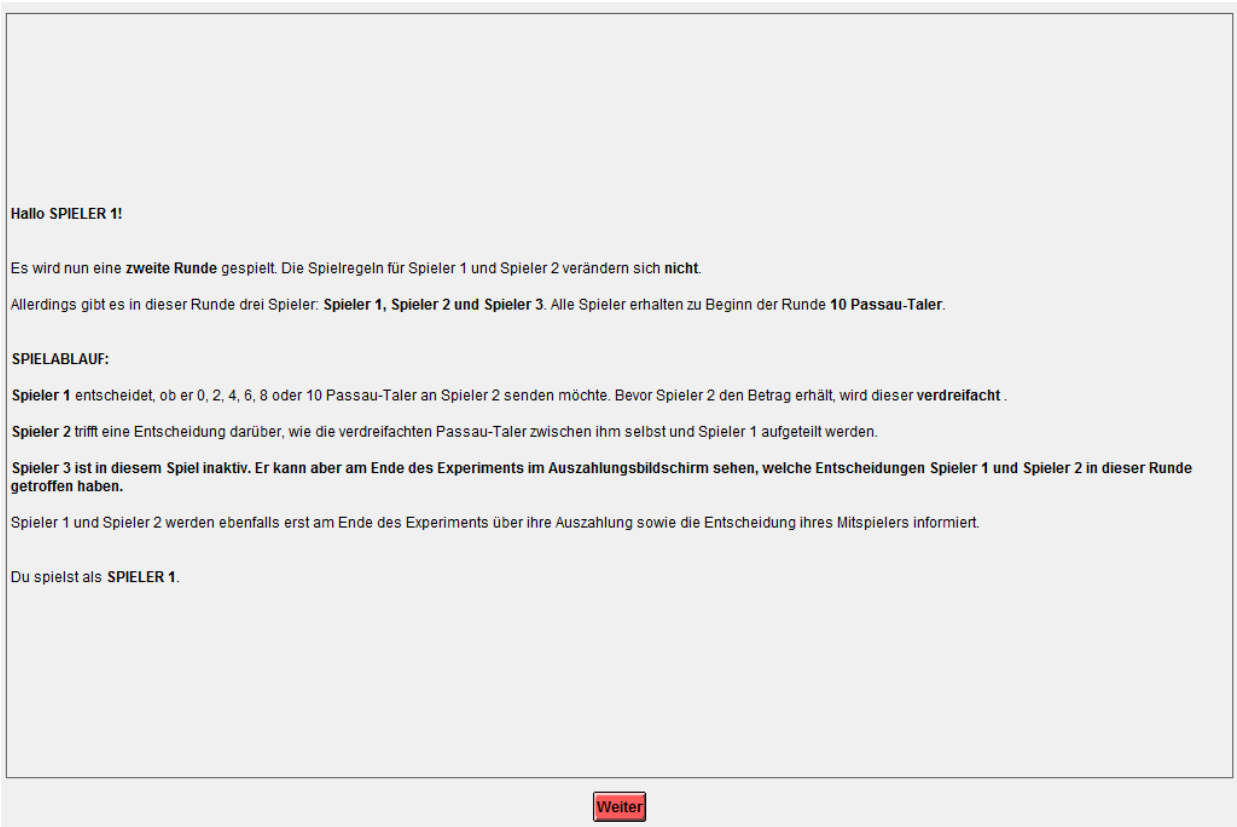
Screenshot 2: Instruction Treatment 1 (Trustor)
(Similarly for Trustee and Third-Party)



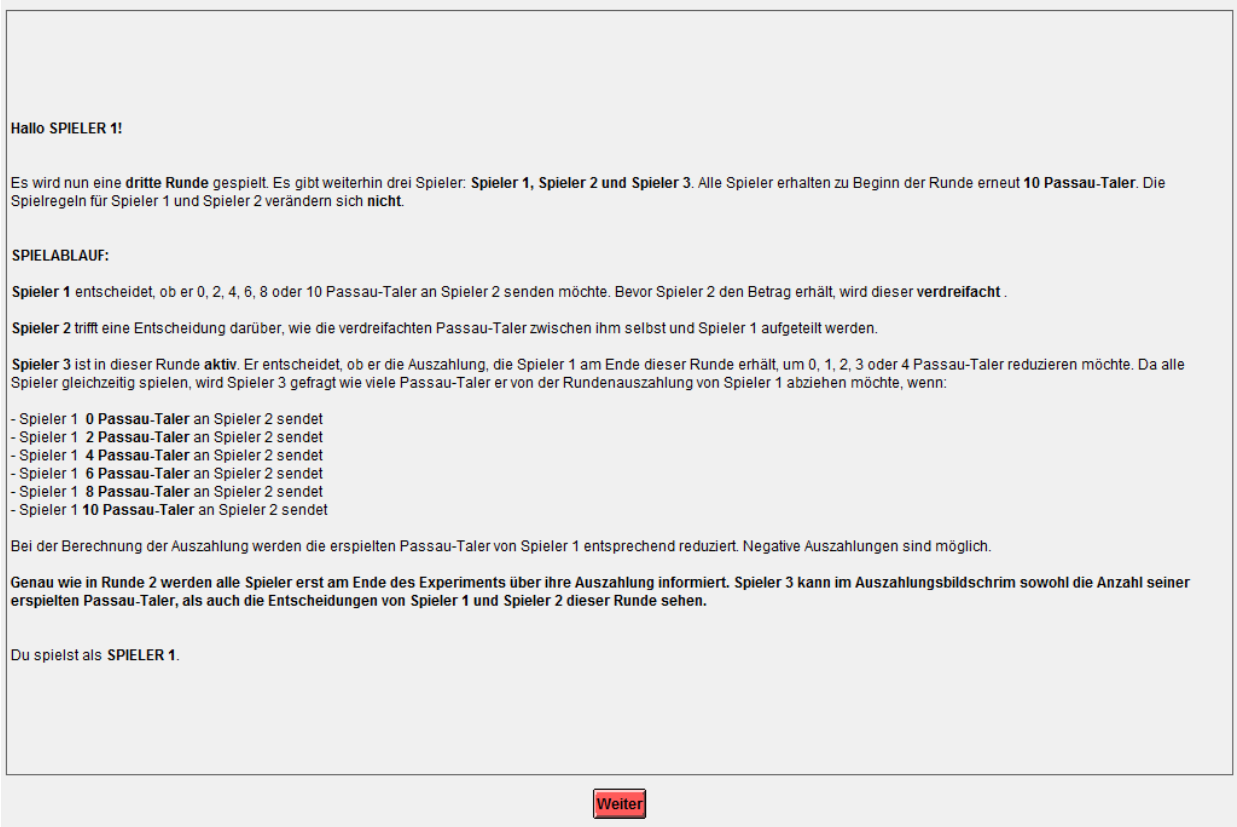
Screenshot 3: Investment Decision (Trustor), Treatment 1
(Similarly for Treatments 2 and 3)



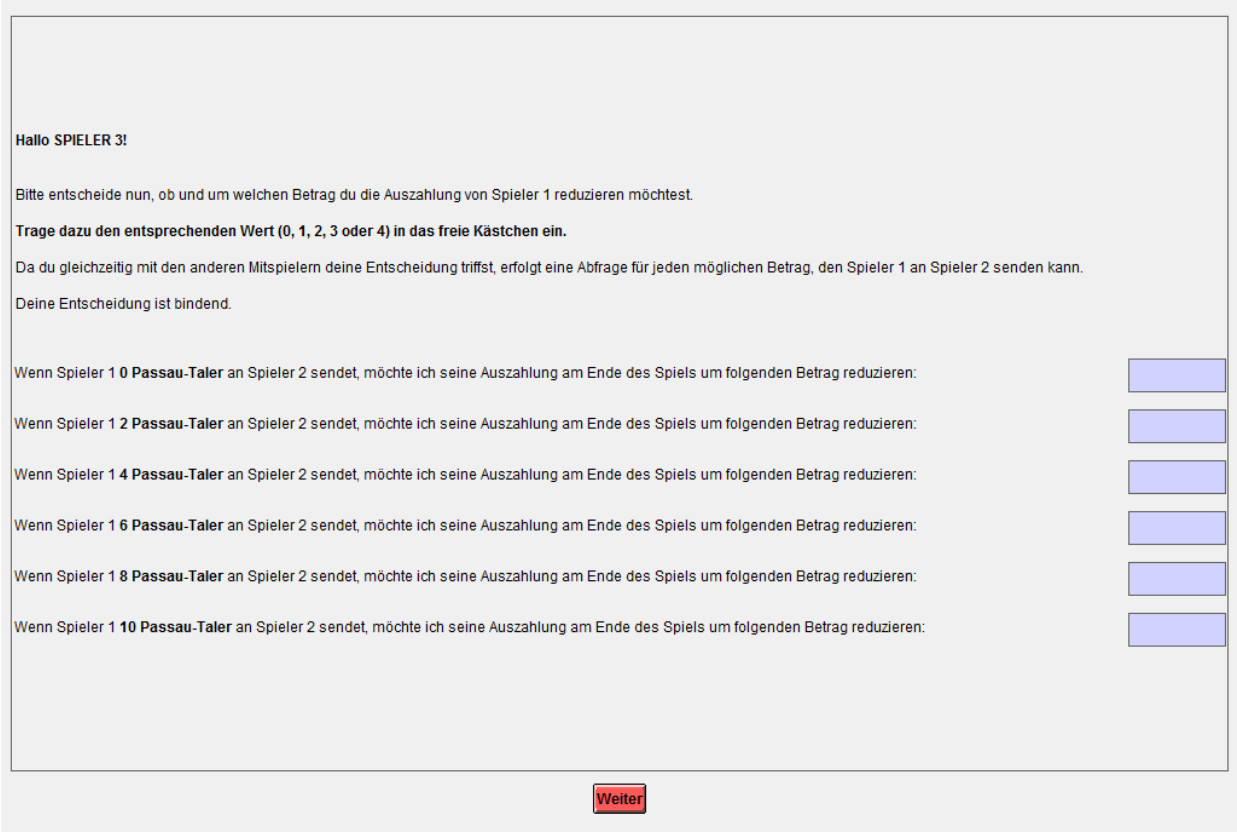
Screenshot 4: Payback Decision (Trustee), Treatment 1
(Similarly for Treatments 2 and 3)



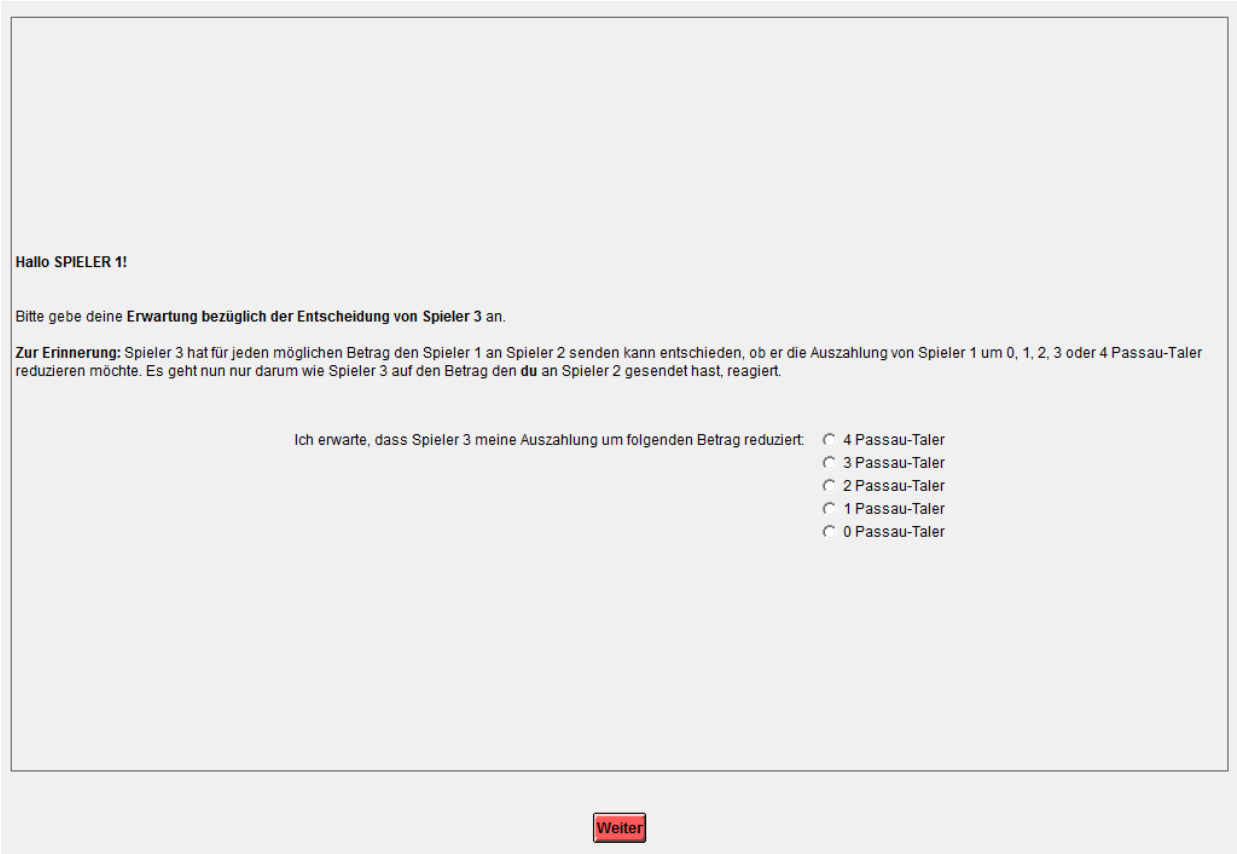
Screenshot 5: Instruction Treatment 2 (Trustor)
(Similarly for Trustee and Third-Party)



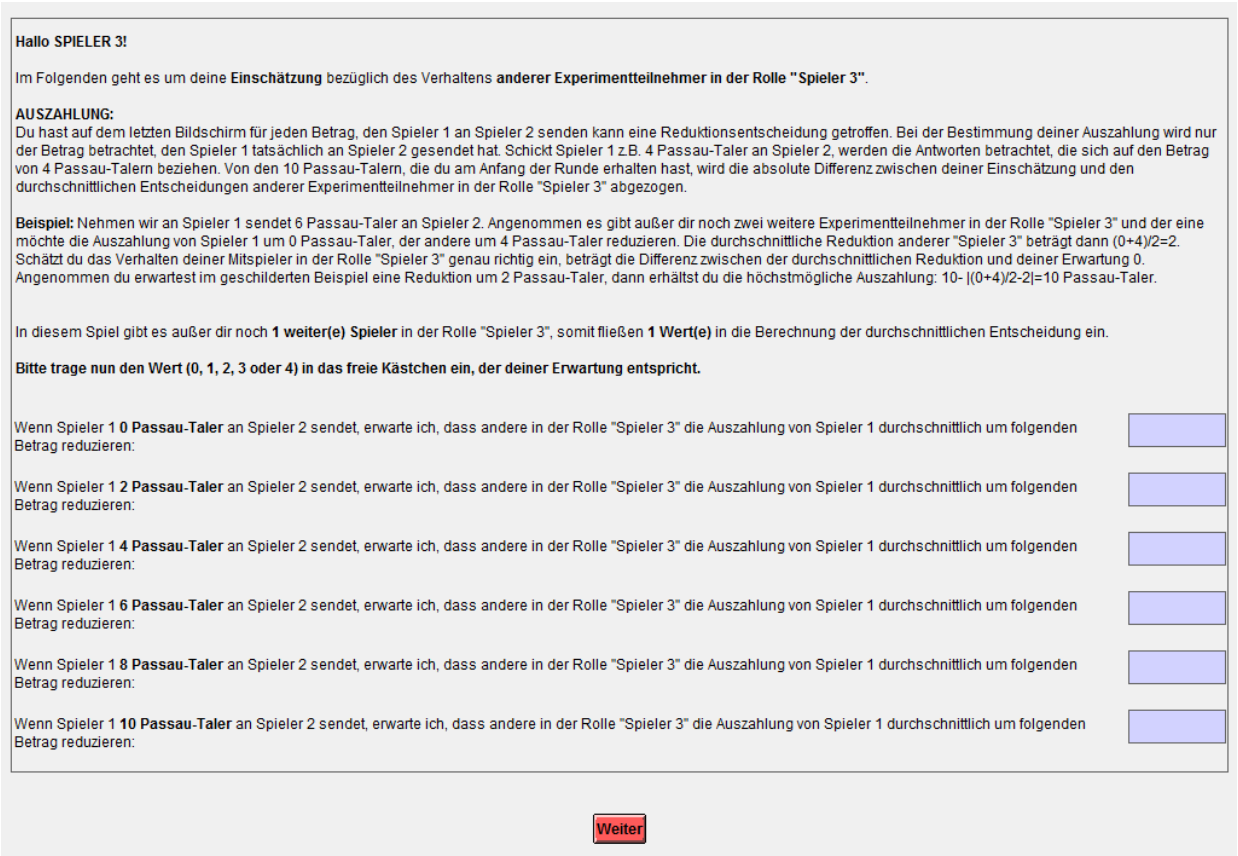
Screenshot 6: Instruction Treatment 3 (Trustor)
(Similarly for Trustee and Third-Party)



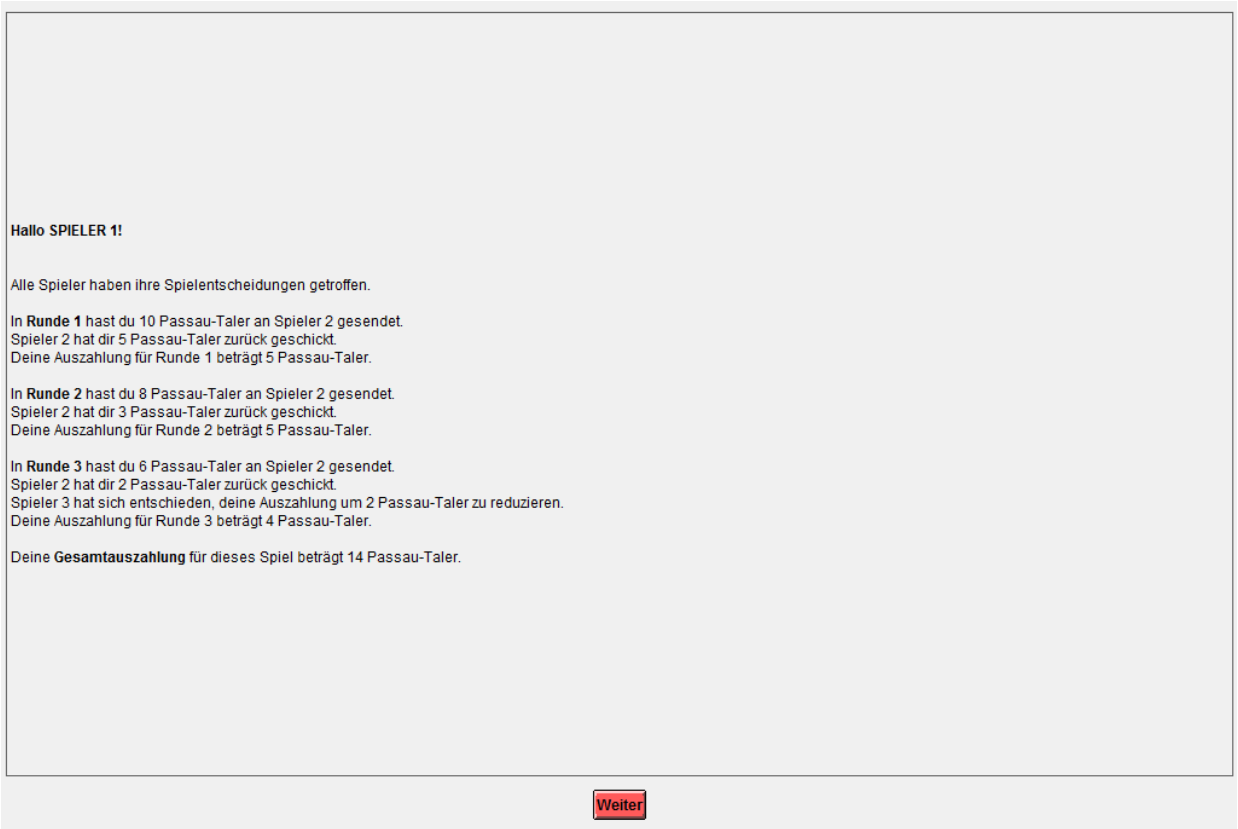
Screenshot 7: Punishment Decision (Third-Party), Treatment 3
(Similarly for Trustee and Third-Party)



Screenshot 8: Punishment Belief (Trustor), Treatment 3



Screenshot 9: Punishment Belief and Payoff-Calculation (Third-Party), Treatment 3



Screenshot 10: Payoff Screen (Trustor)
(Similarly for Third-Party; Trustee Payoff-Screen Shows Averages due to Usage of Strategy Method)