Testing Psychological Forward Induction in the Lost Wallet Game

by

Daniel Woods

A research exercise forming a part of the requirements for the degree

of B.Sc.(Hons)

at the

University of Canterbury

October 2010

Acknowledgements

I would like to thank Maroš Servátka for supervising this Research Project. I would also like to thank the University of Canterbury's Teaching and Learning Committee for funding the experiment, and anyone else who has contributed to this project in any way, shape or form.

Abstract

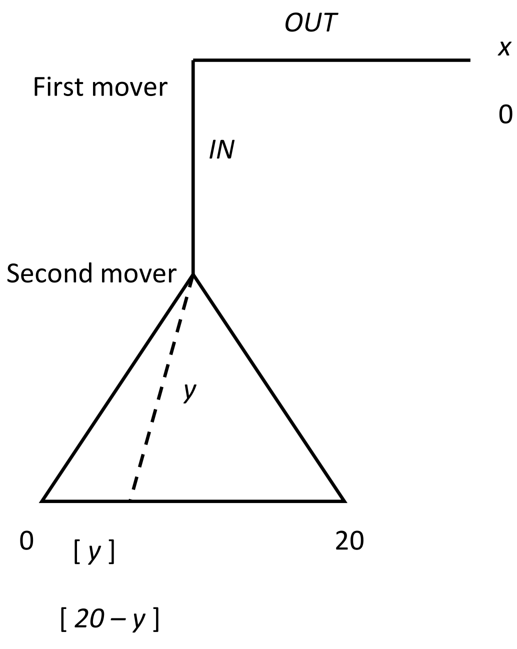
This paper tests Psychological Forward Induction in the lost wallet game, in an attempt to explain an empirical puzzle observed by Dufwenberg & Gneezy (2000). The puzzle observed is that the size of the outside option forgone by the first mover does not affect the behaviour of the second mover. This is puzzling as Psychological Forward Induction and other theories predict that raising the outside option would raise the amount the second mover rewards. This experiment explicitly tests Psychological Forward Induction by eliciting beliefs with different second mover knowledge of first mover decision, depending on treatment. The intention is to observe if knowledge of the first mover's decision affects the second mover's beliefs. The experiment replicates Dufwenberg & Gneezy (2000), and it is found that second movers do update their beliefs conditional on receiving information on the first mover's action, supporting Psychological Forward Induction.

Table of Contents

1. Introduction

The lost wallet game can be described using the following thought experiment. Suppose you find a wallet on the footpath. There is cash in the wallet, which is of value to you, but there are other things in the wallet, like bank cards, drivers licences and photos, which are of no use to you (assuming the bank cards have PIN numbers and you are not interested in identity theft) but have value to the wallet's owner through replacement costs or sentimental value. You have two options, you can take the cash and put the wallet back on the footpath, or you can take the wallet back to the owner who then decides how to reward your good deed. The lost wallet game models this thought experiment as follows. There are two movers. The first mover (FM) has two options: he can go IN, which means that he forgoes the outside option (x), or he can go OUT, and keep the outside option. If the FM goes IN, then the SM gets to split an amount of twenty between the FM (y), and herself ($20 - y$). If the FM goes OUT, then the SM does not make a decision and receives nothing. The extensive form of this game is shown in Figure 1. By backwards induction, the Subgame Perfect Nash Equilibrium for self-regarding preferences is (OUT , y=0).

FIGURE 1 - The lost wallet game tree



A different prediction can be derived by Psychological Forward Induction (PFI) (Battigalli & Dufwenberg, 2009). Consider the FM's decision, who can either go OUT, which makes him better off by x, or he can go IN, which makes him better off by the amount of money the SM decides to reward him, y. If the self-interested FM goes IN, it must be that he expects to receive at least x. [1] Now consider the SM's decision; which is how much to reward the FM for going IN. Let us assume that the SM exhibits 'guilt aversion' (Battigalli & Dufwenberg, 2009), which is an aversion to letting someone down, or not living up to their expectations. If the SM is sufficiently guilt averse, then she will want to reward the FM the amount that the FM expects. However, the SM does not know what the FM expects to receive, so therefore she must form expectations about this amount. Denoting the FM's belief of y as τ', the SM's belief of τ' as τ'', and an individual's sensitivity to guilt aversion as γ, the SM's utility function is now of the form: $20 - y - \gamma(\max\{ \tau'' - y, 0 \})$.[2] PFI predicts that if the FM goes IN, then τ' must be at least x, and the SM, knowing this, updates her belief τ'' to be at least x as well. If the SM is sufficiently guilt averse, then she would send back τ''. [3]

The motivation for this research comes from the results of Dufwenberg & Gneezy (2000), who run the lost wallet game and elicit beliefs about behaviour in the game. Their results show that τ'' and y are positively correlated, which is consistent with guilt aversion, but that x and y are uncorrelated, which is puzzling. Intuitively, using the framework described above, an increasing x means that the SM's τ'' conditional on the FM going IN should increase, which should in turn lead to a higher y. This paper tests PFI explicitly by verifying whether SMs update their beliefs conditional on knowing the action of the FM. In other words, does the SM realise that the FM going IN means that the FM expects at least x and as a consequence, does the SM adjust her beliefs and behaviour accordingly? Such a test involves comparing τ'' between SMs that have information on the FM's decisions, and those that do not.

---

[1] The FM could have different motivations. FMs could be motivated by other-regarding preferences, such as altruism or inequality aversion. For two surveys on other-regarding preferences, see Camerer (2003) or Cooper & Kagel (2009). Alternatively, the FM could trust that the SM will reciprocate and send back at least x. For separation of various FM motives, see Cox (2004).

[2] Note that the SM still prefers more money to less, but her utility function now includes a negative psychological payoff from letting the FM down.

[3] In this case, sufficient guilt aversion would be $\gamma \geq 1$.

The rest of the paper proceeds as follows. In Section 2 I outline the relevant literature involving the lost wallet game and guilt aversion. Section 3 covers how PFI is to be tested and justifies the experimental design. Section 4 provides more specific information on how experiments were conducted. The findings are presented in Section 5, and Section 6 contains discussion and conclusions.

2. Literature Review

Dufwenberg & Gneezy (2000) were the first to experimentally study the lost wallet game. SMs made their decisions using the strategy method (Selten, 1967). Beliefs were elicited from all subjects after their decisions were made. The SMs were asked to state their belief about the FM's expectation of y, but only considering the FMs who went IN. Dufwenberg & Gneezy (2000) found that; (i), subjects did not act in a way consistent with self-regarding preferences, (ii), that as x increases, more FMs go OUT, (iii), that generally, subjects went IN when they expected to receive at least x back, and (iv), that τ'' and y are positively correlated, supporting guilt aversion. However, the surprising result of their experiment was that x and y were uncorrelated. As discussed before, this is inconsistent with PFI, and also with theories of reciprocity (Dufwenberg & Kirchsteiger (2004), Falk & Fischbacher, (2006)).

One hypothesised reason as to why x and y were not correlated was put forward by Cox, Servátka, & Vadovič (2010). They conjectured that the reason that x and y were found to be uncorrelated was because the outside option x was not salient to the SMs. To increase the saliency, they moved to sequential play as opposed to the strategy method to make the SM's decision 'warmer' (Brandts & Charness, 2000). The idea being that it is easier for the SM to consider the FM's monetary consequences of a decision when she observes his decision, before making her own. Also, instead of using decision forms, paper dollar certificates were passed between paired subjects, to ensure that SMs always knew how much the FM had given up in order to go IN, again in an attempt to increase saliency. Beliefs were not elicited as they are not required to look at the correlation between x and y. Despite the saliency measures, Cox et al. (2010) still found that x and y were uncorrelated. An alternative explanation of why the outside option x is ignored which therefore results in x and y being uncorrelated has been tested by Servátka & Vadovič (2009), who explore changing the relative payoffs of the outside option in the lost wallet game. In particular, SMs might ignore the forgone outside option when making their decision if they perceive the outside option to be unfair or unequal. The two treatments tested had either an unequal outside option (10 ,0), or an equal one (5 , 5). However Servátka & Vadovič (2009) found no significant difference in SM's behaviour.

Charness, Haruvy & Sonsino (2003) ran the lost wallet game varying the social distance between the subjects. Three social distance protocols were implemented: a classroom lab experiment, a paired computer lab experiment with the two labs being in different states in the US, and a worldwide internet experiment. In all three treatments the interaction between subjects was anonymous. Relevant to the current paper, they found some evidence for y being correlated with x, but as the strategy method was used in conjunction with a within-subjects design, it is not clear whether this is an actual effect, or an artifact of the experimental design.[4] Finally, evidence for guilt aversion in favour of other belief dependent explanations was found in Charness & Dufwenberg (2006), who ran a hidden action trust game. A principal could choose whether to employ a worker, the worker then decides on either a high or low level of costly effort. If high effort is chosen, a dice roll determines the state, with the bad state having the same payout for the principal as low effort would, hence the choice being hidden to the principal. Beliefs were elicited after decisions were made. The more likely the worker believed that the principal expected high effort, the more likely the worker would put in high effort, which is consistent with guilt aversion.

3. Experimental Design

This experiment consists of two treatments implemented in a between-subjects design. In both treatments, subjects play the lost wallet game presented in Figure 1. The FM makes a decision on whether to go IN or OUT. Going IN forgoes the outside option x, but creates an amount of twenty that is split by the SM between the pair. This results in the monetary payouts of (y , 20-y). Going OUT yields x for the FM, while the SM receives nothing and makes no decisions. In the current experiment subjects play the game sequentially, rather than using the strategy method as in Dufwenberg & Gneezy (2000), to allow for the observability of moves prior to belief elicitation.

[4] Brandts & Charness (2009) survey the literature on the effect of using the strategy method in economics experiments.

The two treatments are termed the information treatment (INFO) and the no information treatment (NO INFO). The treatments differ in when the SM's belief is elicited. In both treatments, the FM is asked to state their belief on the average of *y* before making their decision. In the INFO treatment, the SMs are informed about whether their paired FM has gone IN or OUT, and are asked, considering only FMs that made the same decision, to state their belief of the FMs' belief of the average *y* that was elicited previously. In other words, if a SMs paired FM goes IN, we ask the SM to guess the average FM expectation of *y*, considering only the FMs that went IN, and vice versa for OUT. In the NO INFO treatment, the SMs are not informed of their paired FM's decision, and are asked to state their belief on the average belief of *y* of all FMs, not a subset of FMs as in the INFO treatment. Outside options of x=4 and x=7 were used for comparability with Dufwenberg & Gneezy (2000). [5]

PFI predicts that if the FM goes IN, then τ' must be at least x, and the SM, knowing this, updates her belief τ'' to be at least x as well. The experiment tests whether the thought process of the SM is consistent with what PFI predicts by comparing the beliefs of SMs elicited from the INFO treatment to beliefs of SMs from the NO INFO treatment. If SMs act consistently with PFI, then SMs who are informed that the FM went IN should believe the FM expects $y \geq x$. Similarly, SMs who are informed that the FM went OUT should believe the FM expects $y \leq x$.

As the test of PFI relies crucially on beliefs, the choice of belief elicitation method is of utmost importance. One way of eliciting beliefs is by asking subjects about their expectations in a non-salient way as has been done in many gift-exchange experiments (for example, Brown, Falk & Fehr, 2004). However, non-incentivised elicitation might yield imprecise measures of subjects' beliefs in an experiment. Gächter & Renner (2010) explore the effectiveness of the linear scoring rule in a public good experiment, and find that beliefs are more accurate when the elicitation process is incentivised via a linear scoring rule, as opposed to not being incentivised at all.[6]

[5] Dufwenberg and Gneezy (2000) also used outside options of x=10, x=13 and x=16, which are not explored in this paper as FMs are unlikely to go IN for these values, making observations impractical to get without the use of the strategy method. The same two outside options were also used by Cox et al (2010).

[6] An example of a linear scoring rule is that for every cent a subject's guess is away from the actual value they were trying to guess, the subject loses a cent.

Eliciting beliefs prior to decisions has the potential to change the way subjects behave, as they may place higher cognitive attention on their own and others strategies. The current experiment requires belief elicitation prior to decisions being made, as opposed to after as in Dufwenberg & Gneezy (2000). Gächter & Renner (2010) found that incentivising the beliefs using a linear scoring rule prior to decision making did change subject behaviour in their public good experiment. It is possible that the belief elicitation prior to decisions might change behaviour in the lost wallet game as well. In order to minimize the possible impact of beliefs elicitation on decisions, I follow the protocol of Dufwenburg & Gneezy (2000) and employ the linear scoring rule, which is arguably more practical to implement in an experimental environment than quadratic or logarithmic scoring rules due to ease of explanation to subjects. Finally, it is important to note that the inability to replicate Dufwenberg & Gneezy's (2000) results would also prevent me from being able to provide a meaningful scientific insight into the empirical puzzle observed in their paper.

The beliefs are elicited as follows. FMs are asked what they believe the average of what SMs will return (*y*) is. SMs are asked what they believe the average of the FM's response to the previous belief elicitation, that is, to guess the average expectation of *y* of the FMs. The SM's belief decision is a 'second order belief', as it is a belief about a belief. Second order beliefs are difficult to explain to subjects, which gives further justification of the use of the simpler linear scoring rule.

In essence, the second order belief is being elicited at a different time between the treatments; before and after the paired FM's decision is revealed. The type of second order belief being elicited also differs slightly. The difference in when the FM's decision is revealed is what is actually being tested; do SMs actually update their second order beliefs conditional on receiving information on the FM's decision? It is not possible to test this without withholding information in one of the treatments. Recall that the INFO treatment asks SMs to state their belief on the average of FMs' beliefs on average amounts returned (*y*), but only considering FMs that chose the same action as the SM's paired FM, and that the NO INFO treatment asks SMs to state their belief on the same average, but this time considering all FMs' decisions. Why is the experiment done this way? First, consider the distributions of the aforementioned beliefs. If the conjectured theory is correct, then there should be three distributions of second order beliefs, one for OUT, one for IN, which would be higher than

OUT, and the last would be some combination of the IN and OUT distributions, depending on how many people the SM guessed went IN and OUT. The last distribution would be somewhere in between the IN and OUT distribution, and will therefore be referred to as the 'middle' distribution. If the middle distribution is just a combination of the IN and OUT distributions, then why should one bother to ask for this distribution? The theory could be supported by comparing the difference in the IN and OUT distributions. The middle distribution would be based on the SM's guesses of the number of FMs that choose IN or OUT, which is not relevant to anything in the theory at all. The key assumption here is that the conjectured theory is correct, for which there is no evidence as the theory is precisely what is being tested. In the NO INFO treatment, SMs do not know the FM's decision, and in the INFO treatment they do, the experiment is designed to test whether SMs use this information as predicted by PFI. The alternatives would be asking the SMs in the NO INFO treatment to state their belief only considering FMs that went either IN or OUT, however this could cause confusion in the SM thinking that their paired person had performed whatever action was selected, despite instructions to the contrary. Or, SMs could be asked to state their belief on both types of averages, however it would not be clear if any differences found would be due to use of the within-subject design or the conjectured effect.

4. Experimental Procedures

The experimental sessions were conducted in the New Zealand Experimental Economics Laboratory at the University of Canterbury. Subjects were University of Canterbury students that were recruited through posters and in classes, who then signed up for individual sessions using the online database system ORSEE (Greiner, 2003). The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007). [7] There were ten sessions done overall, with four sessions done with the NO INFO treatment and six sessions done with the INFO treatment. More sessions were required for the INFO treatment to ensure a sufficient number of observations. [8] 190 subjects participated overall, with 112 subjects participating in

[7] Z-Tree is a ready-made experimental software programming environment. The treatments run in sessions were coded from scratch by myself. The z-Tree code for the INFO treatment is included in Appendix C.

[8] This is because observations of τ'' for those going IN and OUT in the INFO treatment are considered separately in the analysis, whereas the NO INFO treatment those observations were pooled.

the INFO treatment, and 78 subjects participating in the NO INFO treatment. The number of subjects in a session varied from fourteen to twenty-four. The sessions were approximately forty-five minutes long, and on average subjects earned $15.60 NZD per session. [9]

In a session, subjects were asked to sign in and then sit quietly until the experiment started. They were asked to sign a consent form, which was collected and then neutrally framed instructions (attached in Appendix B) were handed out, as well as projected onto a screen. Subjects had some time to read the instructions themselves, and then the instructions were read aloud. The same set of instructions were handed out and read to both types of subjects before they were aware of their type (FM or SM), to ensure common knowledge which is especially important in this experiment as SMs are requested to state a second order belief. By not knowing their type while reading the instructions, subjects should pay equal attention to the instructions for both types, and understand the complex second order belief better. Any questions arising were answered in private. Subjects entered their decisions electronically, and upon the completion of the experiment, they were asked to fill out a questionnaire. Subjects were then called one by one to receive their payment in private.

5. Results

First, I test the data on behaviour against Dufwenberg & Gneezy (2000), in order to establish if the change in timing of the belief elicitation affected behaviour. Table 1 reports two tailed p-values from the Mann-Whitney test for differences in distributions of subject's decisions.

TABLE 1 - Mann-Whitney tests on differences in behaviour compared to Dufwenberg & Gneezy (2000)

My data vs. D&G (2000)	Outside option x=4	Outside option x=7	All data
Frequency of going IN	p=.214	p=.513	p=.816
SM's choices of y	p=.449	p=.507	p=.996

[9] For comparison, the minimum wage at the time was $12.75 NZD per hour.

As the lowest p-value is .214, I cannot reject that the distributions of decisions in this experiment are the same as the distributions in Dufwenberg & Gneezy (2000). That is, the distributions of behaviour between the two experiments are likely to be the same. Hence, it is not likely that the different procedures have significantly changed subject's behaviour in this experiment.

I now seek to replicate the main results of Dufwenberg & Gneezy (2000). I first consider their findings on the FMs. Panel 1 in Appendix A contains raw data collected on FM's beliefs and decisions. The first hypothesis tested is that subjects act consistently with what the prediction for self-regarding preferences, for FMs the hypothesis is that they all go OUT.

H1: All FMs choose OUT.

H1 is rejected by inspection, with 69/95, or 73% of FMs going IN. This is consistent with Dufwenberg & Gneezy (2000).

The second hypothesis tested is that FMs go IN if they expect to receive at least *x*. This is one of the steps posited in the process of PFI, and is therefore important to check if it is the case.

H2: FM chooses IN only if his expectation of y is at least x.

There is strong evidence for this hypothesis, with only 3/69 (4%) violations. I also investigate the inverse of this hypothesis, which is that FMs go OUT if they expect to receive less than *x*.

H2': FM chooses OUT only if his expectation of y is less than x.

When *x*=4, all observations violate this hypothesis.[10] When *x*=7, 6/21 (29%) observations violate this hypothesis. It is not as clear whether this hypothesis is supported, as violations of this hypothesis could be due to hedging. [11]

[10] The reason that *x*=4 always violates the hypothesis is possibly due to the linear scoring rule's bias away from boundary beliefs. Consider a simple example; after an exam, a student is asked to guess what their grade was, and if their guess is within five marks of the actual grade, then they will get a reward. The student is somewhat confident they got 100/100 marks but is not entirely sure. If the student guesses 100, the range of marks that he will be rewarded for is 95-100. If the student guesses 95 however, the range is 90-100, so why would the student guess 100 instead of 95? A similar reasoning can be

Now I focus on the hypotheses related to SMs. Panel 2 in Appendix A contains raw data collected on SM's beliefs and decisions. I test for the hypothesis predicted by self-interested preferences, which for SMs is that y=0.

H3: All SMs choose y=0.

55/69, or 80% of SMs that got to make a split, chose a y that was different than zero, so this hypothesis can clearly be rejected by inspection.

I now explore the hypothesis that SM's second order beliefs are correlated with the amount they send back. This hypothesis is related to guilt aversion, if the SM expects that the FM expects to receive more, then the SM will return more to avoid feeling guilty.

H4: y is correlated with τ''.

The Spearman's rank correlation test reports a 2-tailed p-value of p=.0002 if all observations are included and p=.00007 if only x=7 observations are included, showing that it is very likely these variables are correlated. This provides strong evidence for H4, which is consistent with the result found in Dufwenberg & Gneezy (2000). The correlation coefficient is .433 if all observations are included and 0.643 if only the observations from the x=7 treatment is included. This is compared to .4 as reported in Dufwenberg & Gneezy (2000), which included observations over many values of x.

The lack of correlation between x and y is the most interesting result from Dufwenberg & Gneezy (2000). PFI and guilt aversion predicts x and y are correlated. If the SM updates her belief to be more than x if the FM goes IN and then avoids guilt by returning y close to her (updated to be higher) belief, then x and y are correlated.

H5: x and y are correlated.

---

applied for the linear scoring rule, except that rewards decay linearly so the bias is not as extreme. It is expected of FMs that go OUT to guess y=(0-4), but due to the bias, if they actually believe that and they are somewhat risk averse, then they have the incentive to guess approximately y=(4-5), which is above x. This problem is reduced for the x=7 case, as subjects are guessing around a more interior point. It is expected of subjects to go OUT when x=7 if their belief is in the y=(0-7) range, which with the bias is approximately y=(4-7), still below x. There is reason to be suspicious of observations from the x=4 treatments, and so further tests will be run with both all observations, and x=4 or 7 observations separately.

[11] If SMs are more generous than going OUT implies, the FM gets the outside option and a belief accuracy payment, if SMs are not generous, the FM still gets the outside option. If SMs are not generous, the FM still gets the outside option.

The results replicate the puzzling finding of Dufwenberg & Gneezy (2000), with a rejection of H5. The Wilcoxon signed-rank test reports a 2-tailed p-value of .157, so it cannot be rejected that the *y* decisions came from the same distribution when the outside option differed. That is, the *y* distributions from different values of *x* are not likely to be different, suggesting that *x* and *y* are not correlated. The Spearman's rank correlation test reports a p-value of the correlation between *x* and *y* to be p=.515, so the null hypothesis of no correlation cannot be rejected, suggesting also that *x* and *y* are not correlated.

As the findings have replicated conclusions in Dufwenberg & Gneezy (2000), and because I was unable to reject that the distributions of subject's behaviour were the same between the two experiments, I posit that the belief elicitation before the decision has not affected subject behaviour significantly in this game.

Now the hypotheses that this experiment was designed for will be explored. Do SMs update their second order beliefs, given that the SM has received information on her paired FM's decision?

H6: τ'' of SMs that are informed their FM went IN exceeds the τ'' of SMs that are informed their FM went OUT.

H7: τ'' of SMs that are informed their FM went IN exceeds the τ'' of SMs that are not informed of their FM's decision.

H8: τ'' of SMs that are not informed of their FM's decision exceeds the τ'' of SMs that are informed their FM went OUT.

Table 2 reports one tailed p-values from the Mann-Whitney test for differences in distributions. As the theory has a clear prediction for the relative ranking of the distributions of *τ''*, one tailed p-values are appropriate.

TABLE 2 - Mann-Whitney tests on differences in SM's beliefs between treatments

Distributions	x=4	x=7	ALL
IN>OUT	p=.446[12]	p=.000	p=0.001
IN>NO INFO	p=.149	p=.005	p=0.276
NO INFO>OUT	p=.432[11]	p=.108	p=0.015

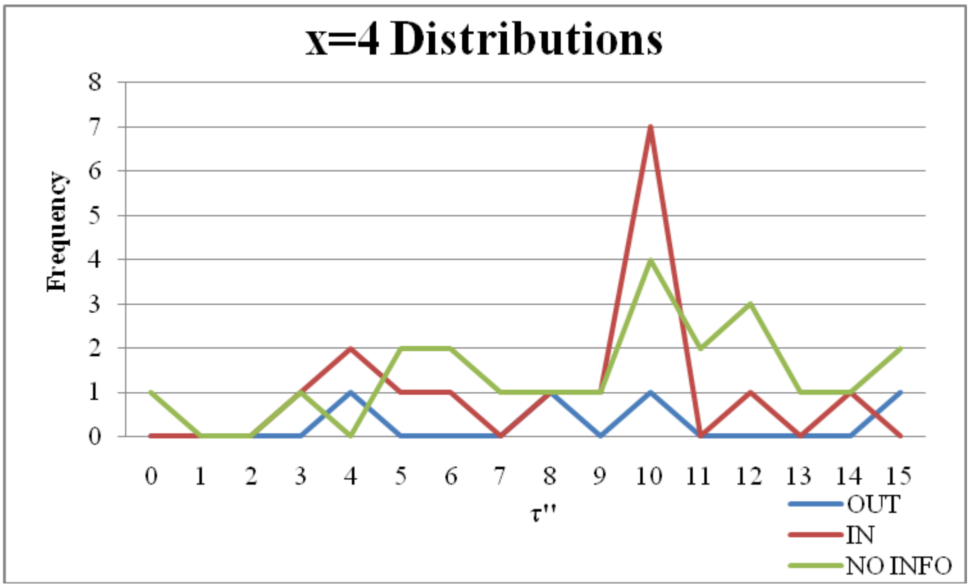
The distributions are the frequencies of τ'' from subjects in the various subsets in which it is posited that there should be differences.

Consistent with H6, the distribution of the second order beliefs of those SMs that were informed that their FM had chosen IN was significantly higher than the distribution of those that were informed that their FM went OUT. SMs are accurately predicting that FMs that go IN are expecting a higher y than those that go OUT.

H7 and H8 are not as clear cut. Note that the IN distribution is clearly higher than the NO INFO distribution when considering beliefs from the x=7 sessions, but that the NO INFO distribution is not statistically significantly higher than the OUT distribution. However, when x=4 observations are included, these results reverse. With clear evidence for H6, it is now appropriate to consider the NO INFO distribution to be a combination of the IN and OUT distributions, which would depend on the proportion of FM actions guessed by the SM. I propose that the reversal of significance between the distributions is due to SMs guessing that a higher proportion of FMs will go IN when x=4, and overwhelmingly so. This would move the middle NO INFO distribution towards the IN distribution and away from the OUT, hence the reversal of significance in the differences. A graphical analysis of the distributions follows.

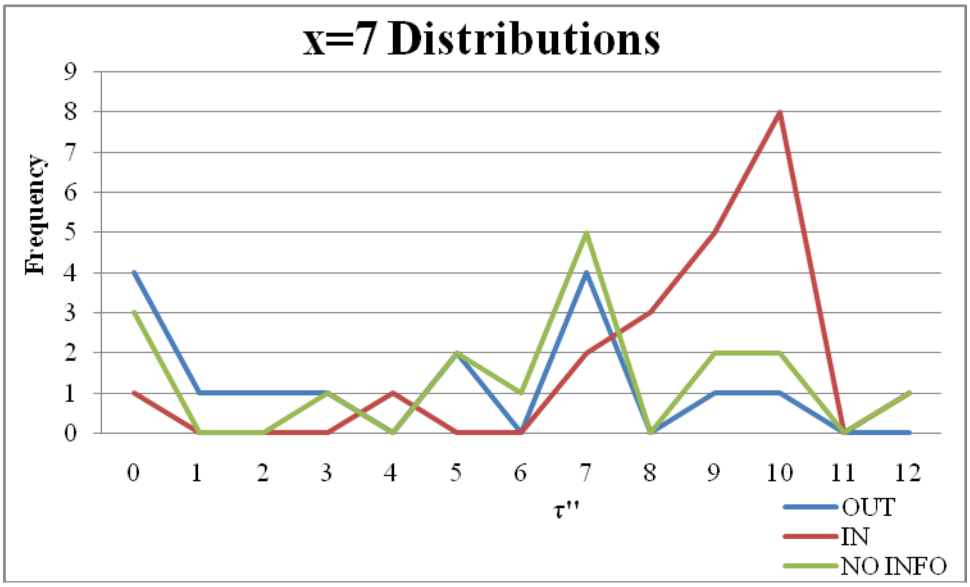
[12] x=4 only had four observations on the OUT distribution, and therefore insignificant results are to be expected.

FIGURE 2 - SM belief distributions when x=4



From inspection, it can be seen that the IN distribution and the NO INFO distribution are somewhat similar. [13] This suggests that SMs in the NO INFO treatment are estimating a high proportion of IN when *x*=4. With only four observations, not a lot can be inferred about the OUT distribution in this situation.

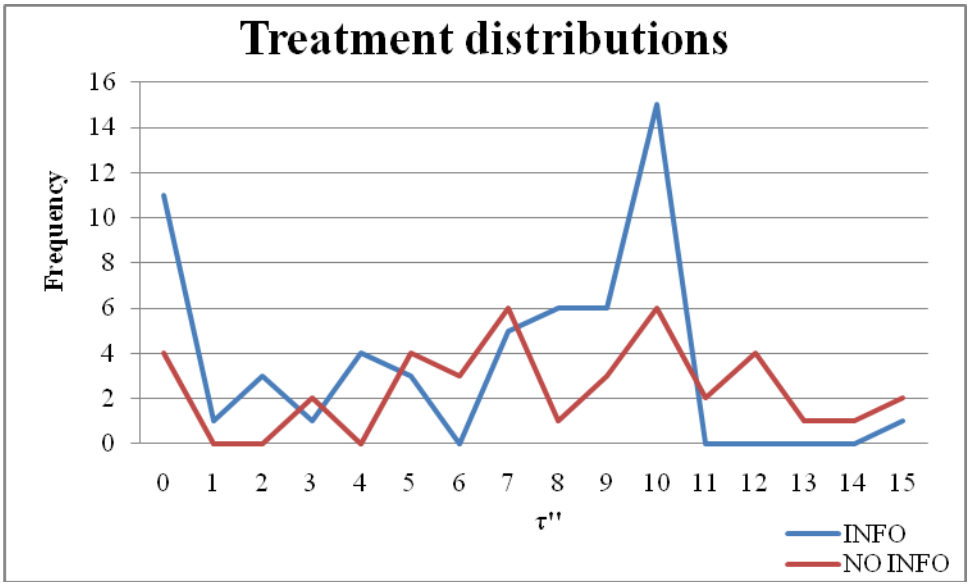
FIGURE 3 - SM belief distributions when *x*=7



[13] Although this is of borderline significance (p=.149), visually the distributions are similar.

Here the OUT distribution seems very similar to the NO INFO distribution, suggesting that subjects in the NO INFO treatment are estimating a high proportion of OUT when *x*=7. The IN distribution is clearly different and higher than the other two distributions.

FIGURE 4 - SM belief distributions pooled by treatment



In the INFO treatment, the distribution seems to be bimodal on zero and approximately nine. This is predicted by PFI, those SMs with a second order belief of zero are likely to have recieved information that the FM had gone OUT. Similarly, SMs with a second order belief of approximately nine are likely to have recieved information that the FM had gone IN . No such behaviour is observed in the NO INFO treatment.

6. Discussion

The results from the experiment suggest that SMs do update their beliefs when given information about the FM's decision. This means that SMs realise that if a FM goes IN, then they expect to recieve more than if they went OUT, and update their second order beliefs accordingly. However, it still holds that the outside option and the amount returned are uncorrelated, but this is not due to a failure to update beliefs. The missing link in the PFI chain that has not been tested is the assumption that guilt aversion drives SM's behaviour. It is true that the SM's second order beliefs are correlated with the amount returned, which would initially suggest that guilt aversion is how SMs are behaving, that as their belief of the

FM's expectation of the amount returned is increasing, they return more to avoid guilt. However, this is not the only explanation that could explain this correlation.

Ellingsen, Johannesson, Tjøtta & Torvisk, (2010) propose that what is actually occurring in games that appear to be demonstrating guilt aversion is the consensus effect. SMs might think that their own belief is shared by the rest of the population, whether this is correct or not. If the SMs think that everyone else would split the money the same as they would, then their stated belief and the amount returned would be the same, hence the correlation between beliefs and amounts returned. Ellingsen et al. (2008) run the lost wallet experiment (along with other games that might exhibit guilt aversion), with the main difference in procedure being that instead of asking the SM for her belief of the FM's expectation, she is given the FM's belief. The FM did not know that the SM would receive this belief, removing any strategic biases, and the belief was elicited via an incentivised scoring rule. Under this treatment the correlation between the second order belief and the amount returned disappears, except when the analysis is limited to modal beliefs, suggesting that guilt aversion might not be what is influencing SM's behaviour in the lost wallet game.

The lost wallet game is 'special', in that the surplus created by investment is always constant at 20, regardless of the outside option forgone. This is different than the well known 'BDM' investment game (Berg, Dickhaut, & McCabe, 1995), in which the amount invested is multiplied by some factor. Cox, Friedman & Gjerstad (2007) examine reciprocity in BDM type investment games (among other games), finding a correlation between variables analogous to the outside option forgone and the amount returned. This suggests that SMs may only be focusing on the amount of surplus generated, and rewarding the FM based on this (Cox, Friedman & Sadiraj, 2008). As the surplus is fixed in the lost wallet game, this could explain why correlation between the outside option and the amount returned is not observed. Therefore, based on the results from this experiment, and in light of the previous literature, I conclude that the puzzling result from Dufwenberg & Gneezy (2000) of x and y not being correlated in the lost wallet game is not due to a failure of SMs to update their beliefs. It is likely an artifact of the game itself, with the surplus being fixed at twenty, or a failure of the assumption of guilt aversion to describe SM behaviour.

As to whether one, the other, or both of the posited effects causes the lack of correlation between the outside option and the amount returned remains an open question that

I leave to future research. Ideally, the experimental design would separate the two effects, and compare the results to the combination of the effects. As a side note, this paper is not about the situation of losing a wallet and what one should do if they found a wallet, despite the analogies used to introduce the game.[14] The situation is an interesting one, but there are various confounds introduced by considering the decisions about the lost wallet situation, including social expectations, entitlement effects of the SM over the wallet, and so forth. These are potentially important factors that could influence behaviour. Therefore, adding context to the current game might change the results. As before, I leave this to future research.

[14] If the reader is interested in a field experiment related to lost wallets, they are invited to read West (2003), who does so.

7. References

Battigalli, P., & Dufwenberg, M. (2009). Dynamic Psychological Games. *Journal of Economic Theory, 144*, 1-35.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust Reciprocity and Social History. *Games and Economic Behaviour, 10*, 122-142.

Brandts, J., & Charness, G. (2000). Hot vs. Cold: Sequential Responses and Preference Stability in Experimental Games. *Experimental Economics, 2*, 227-238.

Brandts, J., Charness, G. (2009). The strategy method: a survey of experimental evidence. Mimeo, Institut d'Anàlisi Econòmica, CSIC, Bellaterra, Spain

Brown, M., Falk, A., & Fehr, E. (2004). Relational Contracts and the Nature of Market Interactions. *Econometrica, 72* (3), 747-780.

Camerer, C.F. (2003). *Behavioral Game Theory: Experiments on Strategic Interaction*. Princeton: Princeton University Press.

Charness, G., Dufwenberg, M. (2006). Promises & partnership. Working papers in Economics, Stockholm University, Sweden.

Charness, G., Friedman, D., & Gjerstad, S. (2007). A tractable model of reciprocity and fairness. *Games and Economic Behaviour, 59,* 17-45.

Charness, G., Haruvy, E., & Sonsino, D. (2000). Social Distance and Reciprocity: The Internet vs. the Laboratory. Mimeo, Technion, Israel.

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

Cox, J.C. (2007). A tractable model of reciprocity and fairness. *Games and Economic Behaviour, 59*, 17-45.

Cox, J. C., Friedman, D., & Sadiraj, V. (2008). Revealed Altruism. *Econometrica, 76* (1), 31-69.

Cox, J. C., Servátka, M., & Vadovič, R. (2010). Saliency of outside options in the lost wallet game. *Experimental Economics, 13,* 66-74.

Dufwenberg, M., & Gneezy, U. (2000). Measuring Beliefs in an Expermental Lost Wallet Game. *Games and Economic Behaviour, 30*, 163-182.

Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behaviour, 47*, 268-298.

Ellingsen, T., Johannesson, M., Tjøtta, S., & Torvisk, G. (2010). Testing guilt aversion. *Games and Economic Behaviour, 68*, 95-107.

Falk, A., & Fischbacher, U. (2006). A Theory of Reciprocity. *Games and Economic Behaviour, 54*, 293-315.

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

Greiner, B. (2003). An Online Recruitment System for Economic Experiments. In: K. Kremer, V. Macho, (Eds.), *Forschung und wissenschaftliches Rechnen 2003* (pp. 79-93). GWDG Bericht 63, Göttingen: Ges. für Wiss. Datenverarbeitung.

Selten, R., (1967). Die Strategiemethode zur Erforschung des eingeschrankt rationalen Verhaltens im Rahmen eines Oligopolexperimentes. In: H. Sauermann, (Ed.), *Beitrage zur experimentellen Wirtschaftsforschung* (pp. 136-168). J.C.B. Mohr (Paul Siebeck), Tubingen.

Servátka, M., & Vadovič, R. (2009). Unequal outside options in the lost wallet game. *Economics Bulletin, 29* (4), 2870-2883.

West, M. D. (2003). Losers: Recovering Lost Property in Japan and the United States. *Law & Society Review, 37* (2), 369-424.

APPENDIX A - Data collected

PANEL 1

INFO Treatment, x=4

IN	τ'
0	5
0	6
0	8
0	12
1	5
1	5
1	5
1	5
1	6.5
1	7.5
1	8
1	8
1	9
1	9
1	10
1	10
1	10
1	10
1	10
1	10

INFO Treatment, x=7

IN	τ'
0	0
0	2.5
0	5
0	7
0	7
0	7
0	7
0	7
0	7
0	7.5
0	8
0	9
0	10
0	13
0	18
1	0
1	5
1	7
1	7
1	7
1	7.5
1	8
1	8
1	8
1	8.5
1	9
1	9
1	9
1	10
1	10
1	10
1	10
1	10
1	10
1	10
1	10

NO INFO treatment, x=4

IN	τ'
0	10
1	1
1	4.4
1	5
1	5
1	6
1	7
1	7
1	7
1	8
1	8
1	8
1	8
1	8
1	8
1	8
1	9
1	9
1	10
1	10
1	10
1	10

NO INFO treatment, x=7

IN	τ'
0	5
0	5
0	5.5
0	6
0	7
0	7
1	7
1	8.5
1	9
1	10
1	10
1	10
1	10
1	10
1	10
1	10
1	10

PANEL 2

INFO treatment, x=4

IN	y	τ''
0	-	4
0	-	7.5
0	-	10
0	-	15
1	0	4
1	0	4
1	0	10
1	0	14
1	2	5
1	4	3
1	4	8
1	8	9
1	8	10
1	8	12
1	9	9.5
1	10	6
1	10	10
1	10	10
1	10	10
1	10	10

INFO treatment, x=7

IN	y	τ''
0	-	0
0	-	0
0	-	0
0	-	0
0	-	1
0	-	1.8
0	-	3
0	-	5
0	-	5
0	-	7
0	-	7
0	-	7
0	-	7
0	-	9
0	-	10
1	0	7
1	0	7.5
1	0	10
1	2	0
1	4	8
1	5	9
1	7	7
1	8	9
1	8	10
1	9	8
1	9	9
1	9	9
1	9	10
1	10	4
1	10	9
1	10	10
1	10	10
1	10	10
1	10	10
1	10	10
1	10	12

NO INFO treatment, x=4

IN	y	τ''
0	-	3
1	0	0
1	0	12
1	0	15
1	1	14
1	2	5
1	5	5
1	6	6
1	6	6
1	6.5	7
1	7	15
1	8	12
1	10	7.5
1	10	9
1	10	10
1	10	10
1	10	10
1	10	10
1	10	11
1	10	13
1	12	11
1	12	12

NO INFO treatment, x=7

IN	y	τ''
0	-	0
0	-	0
0	-	6.5
0	-	7
0	-	8.5
0	-	12
1	0	3
1	0	5
1	0	6
1	0	7
1	1	0
1	5	5
1	5	7
1	6	7
1	8	9
1	8	10
1	10	10

APPENDIX B - Instructions

INSTRUCTIONS (Outside option = 4)

(Instructions specific to the INFO treatment are contained in [square brackets], while instructions specific to the NO INFO treatment are contained in {curly brackets})

No Talking Allowed
Now that the experiment has begun, we ask that you do not talk. If you have a question after we finish reading the instructions, please raise your hand and the experimenter will approach you and answer your question in private.

Show up Fee

Every participant will get $5 as a show up fee, and in addition you may earn money in the experiment. All the money will be paid to you in cash at the end of the experiment.

Anonymity
You will be divided randomly into two groups, called Group 1 and Group 2. Each person in Group 1 will be anonymously paired with a person in Group 2. No one will learn the identity of the person he/she is paired with.

The structure of the experiment
Each person will participate in two tasks. We first describe the DECISIONS task (referred to as TASK 2 in the software) and then the GUESSING task (TASK 1) that precedes it.

DECISIONS
Each person in Group 1 will start with $4 and will have two options:

- To keep all of the $4. In this case the paired Group 2 person with whom he/she is paired makes no decision.
- To send all of the $4 to the paired person in Group 2. In that case the paired person in Group 2 will get to split $20 between the pair. That is, the person in Group 2 will decide how much of the $20, between $0 and $20, to give to the person in Group 1, and how much to keep.

Group 1 GUESS
Each Group 1 person is asked to guess the average amount, between $0 and $20, that people in Group 2 decide to give to the people in Group 1. You will be rewarded based on the accuracy of your answer. How you are rewarded for this guess is outlined in the next paragraph.

In order to check whether your guess is accurate, the average from the Group 2 people's decisions will be calculated. Note that if a Group 2 person's paired Group 1 person keeps the $4, he/she does not get to make a split, so you should not be considering those people when you make your guess. You will be rewarded in the following way: You will start with $5, and for every 1 cent of mistake, 1 cent will be deducted from this $5. The mistake is the absolute value of your guess minus the actual average. For example, if you will guess accurately, you will get $5. If you miss by, say $2, i.e., your guess is either two dollars too high or two dollars too low, you will be paid $3. If your mistake will be larger than or equal to $5, then you will not be paid at all for this part.

Group 2 GUESS

[Each Group 2 person will first be informed whether his/her paired Group 1 person has sent or kept the $4. Each Group 2 person is then asked to guess what was the average guess of the people in Group 1, **but only considering the Group 1 persons that also chose the same decision as <u>your</u> paired Group 1 person**. That is, if your paired Group 1 person sent the $4, you should only consider the guesses of the Group 1 people that also sent the $4. If your paired Group 1 person kept the $4, you should only consider the guesses of the Group 1 people that also kept the $4. Group 1 persons that did not make the same decision will not be included when calculating the average you are guessing. You will be rewarded based on the accuracy of your answer. How you are rewarded for this guess is outlined in the next paragraph.]

{ Each Group 2 person is asked to guess what was the average guess of all the people in Group 1. You will be rewarded based on the accuracy of your answer. How you are rewarded for this guess is outlined in the next paragraph.}

In order to check whether your guess is accurate, the average from the Group 1 person's guesses on how much Group 2 persons will give will be calculated. You will be rewarded in the following way: You will start with $5, and for every 1 cent of mistake, 1 cent will be deducted from this $5. The mistake is the absolute value of your guess minus the actual average. For example, if you will guess accurately, you will get $5. If you miss by, say $2, i.e., your guess is either two dollars too high or two dollars too low, you will be paid $3. If your mistake will be larger than or equal to $5, then you will not be paid at all for this part.

Payment of Show up Fees and Experiment Earnings

All participants are asked to sit patiently until the end of the experiment. Once all Group 2 persons have made their decisions, you will be presented with a summary screen of your earnings, including the averages you were guessing. Click OK after you have seen this screen, so other participants cannot see your decisions. You will then be prompted to complete a Questionnaire. After the Questionnaire, you will be asked one by one to approach the payment room at the back of the lab for the payment of your earnings. Because your decision is private, we ask that you do not tell anyone your decision or your earnings either during or after the experiment. We also ask you to leave using the stairs and not gather in front of the elevators after you receive your payment.

Are there any questions?

APPENDIX C - Z-Tree code for INFO treatment

Setup

code:

```
Wallet = 20;

x = 4;

Type = 0;
```

```
P1_Guess = 0;

Ave_P1_Guess_In = 0;

Ave_P1_Guess_Out = 0;

P2_Guess = 0;

Ave_Reward = 0

Show_Up_Fee = 5;

In=2;
```

Show Group Stage

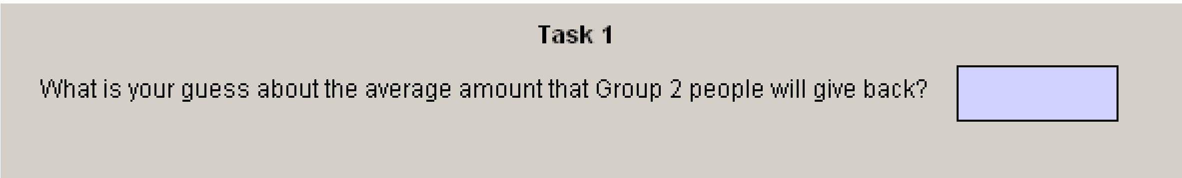
Subject's screen:

P1 Task 1 stage

code:

```
Participate = if ( Type ==1 , 1 , 0 );
```

Subject's screen:

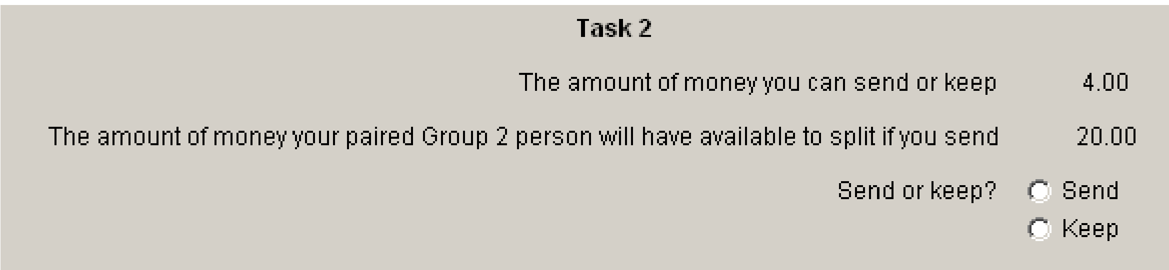


P1 Task 2 stage

code:

```
Participate = if ( Type == 1,1,0);
```

Subject's screen:



P2 Task 1 stage

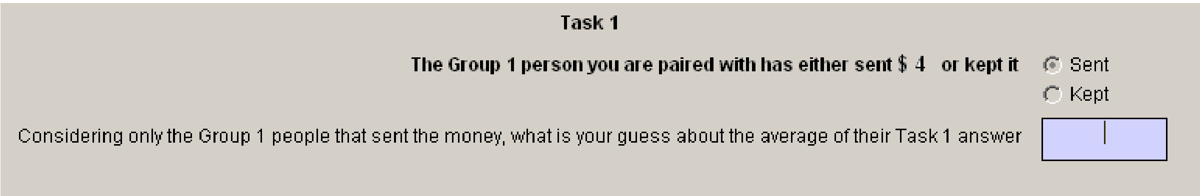
code if IN:

```
if (Type==2) {
     In = find( same(Group) & Type==1, In);
}
Participate = if (Type==2 & In==1, 1,0);
Ave_P1_Guess_In = round( average(not(same(Type)) & In==1, P1_Guess), .01);
```

code if OUT:

```
Participate = if (Type==2 & In==0, 1,0);
Ave_P1_Guess_Out = round( average(not(same(Type)) & In==0, P1_Guess), .01);
```

Subject's screen:



P2 Task 2 stage

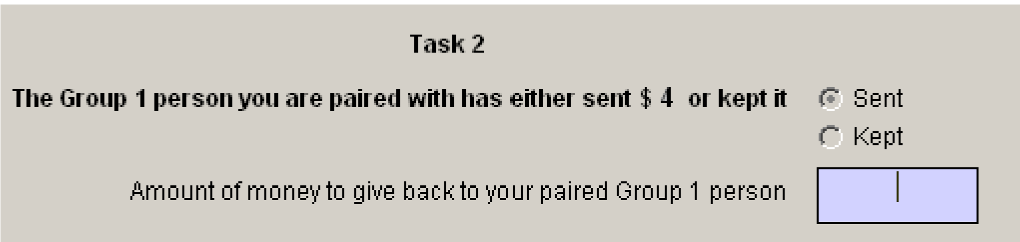
code if IN:

```
Participate = if (Type==2 & In==1, 1,0);
```

code if OUT:

```
Participate = if (Type==2 & In==0, 1,0);
```

Subject's screen if IN:



P1 payout

code if IN:

```
Ave_Reward = round(average(not(same(Type)) & In==1 ,y),.01);
P1_Diff = abs( P1_Guess - Ave_Reward );
if (P1_Diff <= 5) {
     P1_T1_Payout = abs(5 - P1_Diff) ;
}
else {
     P1_T1_Payout = 0;
}
if (In==1) {
     P1_T2_Payout = find(same(Group) & Type==2, y);
}
else {
     P1_T2_Payout = x;
}
if (Type==1 & In==1) {
Profit = P1_T1_Payout + P1_T2_Payout + Show_Up_Fee;
}
```

code if OUT:

```
Participate = if (Type==1 & In==0, 1,0);
p = count (In==1);
if (p==0) {
     P1_T1_Payout = 5;
}
if (Type==1 & In==0) {
Profit = P1_T1_Payout + P1_T2_Payout + Show_Up_Fee;
}
```

Subject’s screen:

Your guess about the average amount of money Group 2 people gave back	5.00
The actual average amount of money Group 2 people gave back was	6.00
Task 1 payout	4.00
You sent the money	
(Task 2 payout) The Group 2 person you were paired with gave back	6.00
Your show-up fee	5.00
Your total payout	15.00

P2 payout

code if IN:

```
Participate = if (Type==2 & In==1, 1,0);
n = count (In==1);
P2_Diff_In = abs(Ave_P1_Guess_In - P2_Guess);
if (P2_Diff_In <= 5 ) {
     P2_T1_Payout_In = abs(5 - P2_Diff_In);
}
else {
     P2_T1_Payout_In = 0;
}
if (n==0) {
     P2_T1_Payout_In = 5;
}
P2_T2_Payout_In = 20-y;
if (Type==2 & In==1) {
Profit = P2_T1_Payout_In + P2_T2_Payout_In + Show_Up_Fee;
}
```

code if OUT:

```
Participate = if (Type==2 & In==0, 1,0);
m = count ( In==0);
P2_Diff_Out = abs(Ave_P1_Guess_Out - P2_Guess);
```

```
if (P2_Diff_Out <= 5 ) {
     P2_T1_Payout_Out = abs(5 - P2_Diff_Out);
}
else {
     P2_T1_Payout_Out = 0;
}
if (m==0) {
     P2_T1_Payout_Out = 5;
}
P2_T2_Payout_Out = 0;
if (Type==2 & In==0) {
Profit = P2_T1_Payout_Out + P2_T2_Payout_Out + Show_Up_Fee;
}
```

Subject's screen:

Of the Group 1 people that sent the money, your guess about their average Task 1 answer was	4.50
Actual average Task 1 answer of Group 1 people that sent the money	5.00
Task 1 payout	4.50
The amount of money that you gave to your paired Group 1 person	6.00
(Task 2 payout) The amount of money you kept for yourself	14.00
Your show-up fee	5.00
Your total payout	23.50