# Testing Psychological Forward Induction and the Updating of Beliefs in the Lost Wallet Game

Woods, Daniel and Servátka, Maroš

Department of Economics, Purdue University, USA, MGSM Experimental Economics Laboratory, Macquarie Graduate School of Management, Sydney, Australia

9 March 2016

# Testing Psychological Forward Induction and the Updating of Beliefs in the Lost Wallet Game

**Daniel Woods[a] and Maroš Servátka[b]**

March 9, 2016

**Abstract**: This paper studies psychological forward induction and the updating of beliefs in the lost wallet game (Dufwenberg & Gneezy, 2000), which is required to derive a prediction for guilt averse agents. Our experiment tests whether the second movers psychologically induct forward and update their beliefs after observing their paired first movers' decision by eliciting beliefs with different second mover knowledge of first mover decision, depending on treatment. We find that second movers do update their beliefs conditional on receiving information on the first mover's action, supporting psychological forward induction.

Keywords: beliefs, experiment, guilt aversion, lost wallet game, psychological forward induction, updating

JEL codes: C70, C91

[a] Department of Economics, Purdue University; email: woodsd42@gmail.com.

[b] Corresponding author. MGSM Experimental Economics Laboratory, Macquarie Graduate School of Management, Sydney, Australia; email: maros.servatka@mgsm.edu.au ; tel. +61-2-98507813.

.

## 1. Introduction

Numerous experiments have shown that when making a decision, people often take into account not only material outcomes but also the expectations of others. Psychological game theory, pioneered by Geanakoplos, Pearce & Stacchetti (1989) and further developed by Battigalli & Dufwenberg (2009), provides a framework that allows incorporating expectations of others and higher order beliefs as part of the payoffs and permits psychological forward induction (henceforth PFI) and the updating of beliefs. The guilt aversion model (Battigalli & Dufwenberg, 2007), according to which the decision-maker forms beliefs about what others expect of her and strives to live up to these (second order) beliefs to avoid feeling guilty, is a prime application of dynamic psychological theory. [1]

Guilt aversion has received considerable attention in the literature, providing insights into motivation of behavior in various strategic environments involving cooperation (e.g. Dufwenberg & Gneezy, 2000; Charness & Dufwenberg, 2006; Bellemare, Sebald & Strobel, 2011). However, the evidence is mixed and subject to methodological issues such as endogeneity of beliefs (Vanberg, 2008) or the false consensus effect (Ellingsen, Johannesson, Tjøtta & Torvisk, 2010), casting doubts on the empirical relevance of guilt aversion. To address these issues a recent study by Khalmetski (2015) exogenously manipulates beliefs and finds that people might not necessarily want to live up to any expectations of others, but only to those that appear to be legitimate from their own perspective, providing evidence that guilt aversion indeed can be a powerful motivation in certain contexts. Additionally, Khalmetski, Ockenfels & Werner (2015) formulate an extension to the guilt aversion model, where people not only feel guilty about not meeting other's expectations, but feel good about exceeding them. This recent work shows that the jury is still out on guilt aversion making further empirical testing relevant..

[1] In sequential games, where the expectations of the other person are not pinned down explicitly by social norms (e.g. leaving a 15% tip) or communication (e.g. forming an informal agreement as in Dufwenberg, Servátka & Vadovič, 2015), a guilt averse agent can use PFI to update her second order belief based on the observed action of her counterpart.

We contribute to this debate by conducting a direct test of PFI and belief updating in the lost wallet game (Dufwenberg & Gneezy, 2000; henceforth DG), which is required to derive a prediction for guilt averse agents. In this two-player game, the first mover (FM) has two options: he can choose IN, which means that he forgoes the outside option ($x$), or he can choose OUT, and keep the outside option. If the FM chooses IN, then the second mover (SM) gets to split an amount of twenty dollars between the FM ($y$), and herself ($20 - y$). If the FM chooses OUT, then the SM does not make a decision and receives nothing.[2] PFI derives a prediction as follows. Consider the FM's decision, who can either choose OUT, which makes him better off by $x$, or choose IN, which makes him better off by the amount of money the SM decides to reward him, $y$. If the self-interested FM chooses IN, he expects to receive at least $x$.[3] Now consider the SM's decision; which is how much to reward the FM for choosing IN. Let us assume that the SM does not want to let the FM down. If the SM is 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 $\tau'$, the SM's belief of $\tau'$ as $\tau''$, and an individual's sensitivity to guilt as $\gamma$, the SM's utility function is now of the form: $20 - y - \gamma(\max\{\tau'' - y, 0\})$. The SM still prefers more money to less, but her utility function now includes a negative psychological payoff of feeling guilty from letting the FM down. PFI predicts that if the FM chooses IN, then $\tau'$ must be at least $x$, and the SM, knowing this, updates her belief $\tau''$ to be at least $x$ as well. If the SM is sufficiently guilt averse (in this case $\gamma \geq 1$), she will send back $\tau''$.

DG's experimental results show that $\tau''$ and $y$ are positively correlated, which is consistent with guilt aversion and PFI, but that $x$ and $y$ are uncorrelated, which is puzzling. Intuitively, using the guilt aversion framework described above, an exogenous increase in $x$ (DG's treatment variation to test the guilt aversion predictions) means that the SM's $\tau''$ conditional on the FM choosing IN should increase, which should in turn lead to a higher $y$. Given that data from DG shows that $\tau''$ is not correlated with $x$, it is possible that their subjects fail to psychologically induct forward. They may fail to psychologically induct

[2] By backwards induction, the Subgame Perfect Nash Equilibrium for self-regarding preferences is (OUT, $y$=0).

[3] While our main focus is on PFI, the FM could also 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 and SM motives, see Cox (2004).

forward by not updating $\tau''$ to reflect that a FM who chooses IN expects to receive $x$. If SMs do not correctly update $\tau''$ using PFI, then the results from DG can still be consistent with guilt aversion, as the lack of correlation between $x$ and $y$ can be explained by SMs not realizing that a FM choosing IN is a signal of the FM's expectation of $y$ relative to $x$. However, if SMs do correctly update $\tau''$ in line with PFI, then the results from DG provide evidence against the guilt aversion model being an accurate descriptor of SM behavior.

The current paper provides a further and more direct test of PFI by verifying whether SMs update their beliefs conditional on knowing the action of the FM. In other words, we examine whether after observing the FM choosing IN, the SM updates her beliefs that the FM expects at least $x$ and behaves accordingly. Such a test involves comparing $\tau''$ between SMs who have information on the FM's decisions, and those who do not. In contrast, the test of PFI in DG involves measuring the impact of an exogenous variation in $x$ on $\tau''$. Our method thus allows us to verify whether subjects update their 'within-game' beliefs based on observing a particular action, without this effect potentially interacting with a change in the treatment variable. Decomposing the mechanism through which guilt aversion operates from the assumptions made about preferences enables us to pinpoint reasons why the theory of guilt aversion might fail in some contexts, providing further insights into what drives observed SM behavior.

## 2. Literature Review

DG were the first to experimentally study the lost wallet game. Their SMs made their decisions using the strategy method (Selten, 1967). All subjects had their beliefs elicited after making their decisions. DG asked SMs to state their belief about FM's expectation of $y$, but only considering FMs who chose IN. DG found that; (i) subjects did not act in a way consistent with self-regarding preferences, (ii) as $x$ increased, more FMs chose OUT, (iii) generally, subjects chose IN when they expected to receive at least $x$ back, and (iv), $\tau''$ and $y$ were positively correlated, supporting guilt aversion. However, the surprising result of their experiment was that $x$ and $y$ were uncorrelated. As discussed previously, this is inconsistent with PFI, and also with some 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) who conjectured that this 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 and 2011). The idea is that it is easier for the SM to consider the FM's monetary consequences of a decision when she observes the FM's decision, before making her own. In addition, instead of using decision forms, paired subjects passed paper dollar certificates between them, to ensure that SMs always knew how much the FM had given up in order to choose IN, again in an attempt to increase saliency. Beliefs were not elicited as they were not required to investigate the correlation between *x* and *y*. Despite the saliency measures, Cox et al. (2010) still found that *x* and *y* were uncorrelated. Servátka & Vadovič (2009) tested an alternative explanation of why *x* might be ignored, by exploring changing the relative payoffs of the outside option. 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 implemented two treatments had either an unequal outside option (10, 0), or an equal one (5, 5). However, Servátka & Vadovič found no significant difference in SM's behavior between the two treatments.

Charness, Haruvy & Sonsino (2007) ran the lost wallet game varying the social distance between the subjects. They implemented three social distance protocols: 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 correlation, or an artifact of the design.

Ellingsen et al. (2010) proposed that what is actually occurring in the lost wallet game is the consensus effect. The consensus effect is where individuals believe that their own actions are representative of the general population, which means that the direction of causality between beliefs and actions may be the opposite of what guilt aversion posits. This would explain both the observed correlation between $\tau''$ and *y*, and the lack of correlation between *x* and *y* in the lost wallet game as SMs are estimating $\tau''$ based on *y* instead of *x*. In Ellingsen et al. experiment the SM was informed about the FM's exact belief, which was elicited prior. The FM did not know that the SM would receive this belief, removing any strategic motivations. No correlation between the supplied FM belief (a stand-in for $\tau''$) and $y$ was found, suggesting that SM's may not be guilt averse. However, Khalmetski (2015) pointed out a confound of signaling information about the FM through his belief in Ellingsen et al. design. Khalmetski instead used a game where a random number was generated, which was observed by the FM but not the SM. The FM had to decide whether to tell the truth about the observed number to the SM, and then the SM had to guess the observed number. By changing the incentives for the FM to lie, Khalmetski found that if the incentives were low, FMs were more likely to lie when SMs expected them to do so, consistent with guilt aversion, but observed no difference in lying if the incentives were high, suggesting subjects may be guilt averse but only when incentives are moderate.

Guilt aversion, along with other elements of PFI, therefore does have some empirical support in the literature. A crucial element that remains untested is whether SMs update $\tau''$ in a manner consistent with PFI, which is what we seek to address.

## 3. Experimental Design and Procedures

Our experiment consists of two treatments, the information treatment (INFO) and the no information treatment (NO INFO), implemented in a between-subjects design. In both treatments, subjects play the lost wallet game as described in the introduction. The game is played sequentially, rather than using the strategy method as in DG, to allow for the observability of moves prior to belief elicitation.

The two treatments differ in when we elicit the SM's belief. In both treatments, we ask FMs to state their belief on the average of $y$ before they make their decision. In the INFO treatment, we inform SMs about whether their paired FM chose IN or OUT, and we ask the SM, 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 SM's paired FM chooses IN, we ask the SM to guess the average FM expectation of $y$, considering only FMs that chose IN. If a SM's paired FM chooses OUT, we ask the SM to guess the average FM expectation of $y$, considering only FMs that chose 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. The outside option of $x$=7 was chosen for this experiment to ensure a sufficient number of IN and OUT decisions as the design is not using the strategy method. [4]

PFI predicts that if the FM chooses IN, then $\tau'$ must be at least $x$, and the SM, knowing this, updates her belief $\tau''$ to be at least $x$ as well. The current 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 chose IN should believe the FM expects $y \geq x$. Similarly, SMs who are informed that the FM chose OUT should believe the FM expects $y \leq x$. SMs who are not informed of the FMs decision should report their initial beliefs, which will serve as a baseline to observe if and how SMs update their beliefs.

As the test of PFI relies crucially on beliefs, the choice of belief elicitation method is important. 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 (e.g., Brown, Falk & Fehr, 2004). However, non-incentivized 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 a linear scoring rule incentivizes the elicitation process, as opposed to an unincentivised elicitation process. 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 subjects making their decisions, as opposed to after as in DG. Belief elicitation prior to decisions might affect behavior in the lost wallet game as well, however, in this design it cannot be avoided. In a survey of the literature, Schlag, Tremewan & van der Weele (2015) report mixed empirical evidence on preferred scoring rules and on beliefs affecting behavior, and also discuss the practicality of implementing complex scoring rules. It is with this in mind that we follow the protocol of DG and employ the linear scoring rule. While it is not incentive compatible for the mean if people maximize expected value, the linear scoring rule 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 significantly different procedures

[4] DG also used outside options of x=4, x=10, x=13 and x=16, which are not implemented in the current experiment as varying the size of the outside option is not required to test PFI and the updating of beliefs.

from those implemented in DG's would prevent us from being able to provide a meaningful scientific insight into the empirical puzzle observed in their paper.

The belief elicitation process is as follows. Using a linear scoring rule, we elicit $\tau'$ by asking FMs what they believe the average of what SMs will return is. We also elicit $\tau''$ by asking SMs what they believe the average of the FM's response to the previous belief elicitation. The SM's belief, $\tau''$, 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.

Effectively, we are eliciting SM belief, $\tau''$, at a different time between the treatments; before and after we reveal the paired FM's decision. The type of second order belief we elicit also differs slightly. The difference in when we reveal the FM's decision is what we are testing; do SMs actually 'update' $\tau''$ conditional on receiving information on the FM's decision?[5] 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 amount returned ($y$), but only considering FMs who 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. We designed this experiment in this way in an attempt to elicit a SM's initial beliefs without other potential confounds. An alternative would be asking the SMs in the NO INFO treatment to state their belief only considering FMs that chose either IN or OUT. However, this could cause confusion with SMs thinking that their paired person had performed the presented action, despite instructions to the contrary. If SMs were behaving in such a way, then there would be no difference between our treatments. Alternatively, we could ask SMs to state their belief on both types of averages; however, the presence of both elicitations (a within-subject design) could influence their responses. We believe our approach gives us the most appropriate measure of a SM's initial beliefs, given our available options.

The experiment was conducted in the New Zealand Experimental Economics Laboratory at the University of Canterbury. The participants were selected randomly from the database using the ORSEE recruitment system (Greiner, 2015). The experiment was

[5] Note that we stick to the original terminology of belief updating even though our design is across-subjects.

computerized using the software z-Tree (Fischbacher, 2007). We conducted two sessions with the NO INFO treatment (72 subjects) and three sessions with the INFO treatment (34 subjects). More sessions were required for the INFO treatment to ensure a sufficient number of observations.[6] The sessions were approximately 45 minutes long, and on average subjects earned $15.60 NZD per session.[7]

In each session, we handed out neutrally framed instructions (see Appendix A), and projected them onto a screen. Subjects had some time to read the instructions themselves, and then the experimenter read the instructions aloud. The same set of instructions were handed out and read to both types of subjects before they were made 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. The experimenter answered any arising questions in private. At the end of the session the subjects were paid individually in private.

## 4. Results

We begin by comparing the behavior of our subjects against DG, in order to establish if the change in timing of the belief elicitation affected decisions. Table 1 reports summary statistics and p-values from the Mann-Whitney test for differences in distributions of decisions.[8]

As the statistical tests presented in Table 1 indicate, we cannot reject that the distributions of decisions in this experiment are the same as the distributions in DG (p=.534 and p=.493 for FMs and SMs respectively). Hence, it is not likely that the different procedures have significantly influenced subject behavior in this experiment.

We now seek to replicate some findings from DG. The first hypothesis is that subjects act consistently with what the prediction for self-regarding preferences, for FMs the hypothesis is that they all choose OUT.

---

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

[7] The minimum wage at the time of running the experiment was NZD 12.75 per hour. Subjects could earn up to $5 via the belief elicitation procedure and participate in the lost wallet game as described previously.

[8] Throughout the paper we report two sided p-values. Raw data is reported in Appendix B.

**Table 1**: Behavior of Subjects in Our Experiment and the DG Experiment

**Panel A: First Mover Behavior**

| **Data** | **Fraction of FMs who chose IN** | **Fisher's Exact Test** |
|---|---|---|
| **Our data** | 32/53 (60%) | - |
| **DG x = 7 data** | 6/12 (50%) | - |
| **Our data vs DG x = 7 data** | - | $p$ = .534 |

**Panel B: Second Mover Behavior**

| | **Mean Amount Given to the FM (average *y*)** | **Median Amount Given to the FM (median *y*)** | **Mann-Whitney Test** |
|---|---|---|---|
| **Our data (n = 32)** | 6.03 [4.00] | 8 | - |
| **DG x = 7 data (n = 12)** | 4.83 [4.49] | 4.5 | - |
| **Our data vs. DG x = 7 data** | - | - | $p$ = .493 |

Standard deviations in brackets.

*H1: All FMs choose OUT.*

We reject H1 with 32/53, or 60% of FMs choosing IN, and a one-sample t-test against the hypothesized mean of zero reporting p=.0001. This is consistent with DG.

The second hypothesis is that FMs choose IN if they expect to receive at least $x$. This is one of the steps posited in the process of PFI, and it is therefore important to verify.[9]

[9] Since our focus is on testing PFI and updating of beliefs, outside of the existing design we do not measure subjects' risk aversion, other-regarding preferences, and the dispersion of beliefs, all of which could possibly influence decisions and thus our tests of H2 and H2'. Cf. footnote 3.

*H2: The FM chooses IN only if τ' is at least x.*

We find strong evidence for this hypothesis in our data, with only 2/32 (6%) violations. We also investigate the inverse of this hypothesis, which is that FMs choose OUT if they expect to receive less than *x*.

*H2': The FM chooses OUT only if τ' is less than x.*

We find that 14/21 (67%) observations violate H2'.

Now we focus on SMs. We test the hypothesis predicted by self-interested preferences, which for SMs is that $y = 0$.

*H3: All SMs choose y = 0.*

In our experiment 25/32 (78%) of SMs who got to make a split, chose a *y* that was different than zero. A one-sample t-test against the hypothesized mean of zero reports p=.0001. Consequently, we reject H3.

We now explore the hypothesis that there is correlation between a SM's second order belief and the amount she gives to the FM. 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 p-value of $p$=.00007, providing strong statistical evidence that these variables are correlated. Our finding is consistent with the result found in DG. The correlation coefficient between *y* and *τ''* in our data is .643, compared to .4 in DG data, which included observations over many values of *x*.

As our findings have replicated the applicable conclusions in DG, and because we are unable to reject that the distributions of subject behavior were the same between the two experiments, we posit that the belief elicitation before the decision has not affected subject behavior significantly in this game.

We briefly examine FMs beliefs, as presented in Panel A of Table 2. Our hypothesis is that FMs who choose IN expect on average to receive more from SMs compared to FMs who chose OUT.

*H5: τ' of FMs who choose IN is greater than τ' of FMs who choose OUT.*

H5 is strongly supported by the results of Mann Whitney tests on our pooled data and the NO INFO subset reporting p-values of p=.001 in both cases. The Mann Whitney test on our INFO subset is marginally insignificant with a p-value of .109, however, the averages are in the direction predicted by H5.

Next, we explore the main hypotheses of our experiment. Do SMs update their second order beliefs after receiving information on their paired FM's decision? PFI predicts that SMs will believe that FMs who choose IN expect to receive a higher *y* back than FMs who choose OUT.

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

SMs beliefs are summarized in Panel B of Table 2, and graphically displayed in Figure 1. The Mann-Whitney test comparing the beliefs of SMs whose paired FMs chose IN with the beliefs of SMs whose paired FMs chose OUT reports a p-value of .0004, providing strong evidence for H6. SMs have identified that FMs who choose IN expect to receive a higher *y* than those who choose OUT, and have updated their beliefs accordingly.

We now explore how SMs update their initial beliefs conditional on observing their paired FM's decision. We do this by comparing beliefs in NO INFO treatment (our proxy for initial beliefs) to the relevant subset of our INFO treatment. PFI predicts that SMs will update their initial beliefs upward upon learning their paired FM has chosen IN.

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

From Table 2, Mann-Whitney test comparing the relevant distributions reports a p-value of p=.011, providing strong evidence for H7. SMs are updating their initial beliefs upwards when presented with information about their paired FM choosing IN.

**Table 2:** Subject Beliefs in Our Experiment

**Panel A: First Mover's Beliefs**

| Treatment and Decision | Mean Belief of $y$ (average $\tau'$) | Median Belief of $y$ (median $\tau'$) | Mann Whitney Test |
|---|---|---|---|
| **NO INFO IN** | 9.50 | 10.00 | p=.001 |
| **(n=11)** | (0.97) | | |
| **NO INFO OUT** | 5.92 | 5.75 | |
| **(n=6)** | (0.92) | | |
| **INFO IN** | 8.24 | 9.00 | p=.109 |
| **(n=21)** | (2.35) | | |
| **INFO OUT** | 7.67 | 7.00 | |
| **(n=15)** | (4.12) | | |
| **NO INFO ALL** | 8.24 | 9.00 | p=.703 |
| **(n=17)** | (1.99) | | |
| **INFO ALL** | 8.00 | 8.00 | |
| **(n=36)** | (3.17) | | |
| **ALL IN** | 8.67 | 9.50 | p=.001 |
| **(n=32)** | (2.03) | | |
| **ALL OUT** | 7.17 | 7.00 | |
| **(n=21)** | (3.57) | | |

**Panel B: Second Mover's Beliefs**

| Treatment and Decision | Mean belief of $\tau'$ (average $\tau''$) | Median belief of $\tau'$ (median $\tau''$) | Mann Whitney Test | Mann Whitney Test |
|---|---|---|---|---|
| **NO INFO** | 6.06 | 7.00 | p=.211[a] | p=.011 |
| **(n=17)** | (3.59) | | | |
| **INFO IN** | 8.50 | 9.00 | p=.0004 | |
| **(n=21)** | (2.55) | | | |
| **INFO OUT** | 4.19 | 5.00 | | p=.211[a] |
| **(n=15)** | (3.56) | | | |

[a] denotes the same statistical test comparing NO INFO vs. INFO OUT. Standard deviation in brackets.

Similarly, PFI predicts that SMs will revise their initial beliefs downwards upon learning their paired FM has chosen OUT.

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

From Table 2, there is no strong evidence in support of H8, as the Mann-Whitney test is insignificant reporting a conservative two sided p-value of p=.211. However, the averages are in the hypothesized direction. We cannot say whether SMs are revising their initial beliefs downwards when presented with the information that their paired FM has chosen OUT. Lack of support for H8 would not explain the lost wallet game puzzle, as SMs who observe OUT do not make a *y* decision. With strong evidence for H6 and H7, we conclude that SMs do generally update their beliefs in accordance with the predictions proposed by PFI.

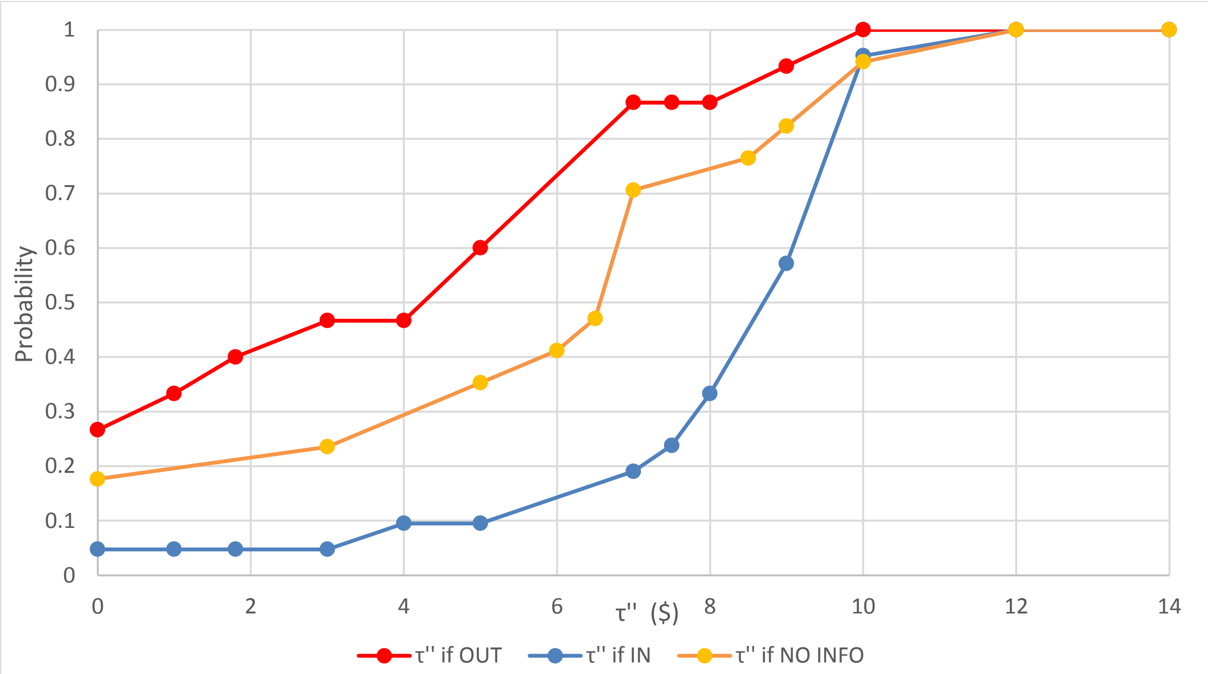


**Figure 1:** Cumulative Distribution Functions of SM beliefs

## 5. Discussion

The results from our experiment suggest that SMs do update their beliefs when they learn about the FM's decision. This means that SMs realize that if FMs choose IN, then such FMs expect to receive a higher *y* than if they chose OUT. As a consequence, SMs update their second order beliefs (*τ''*) accordingly. Our results thus suggest that the robust finding that *x* and *y* are not correlated in the lost wallet game is not due to SMs' failure to update beliefs. Rather, our findings combined with DG's results provide some evidence against guilt aversion. From DG, *τ''* is correlated with *y*, which would initially suggest that guilt aversion drives SMs behavior, i.e., as *τ''* is increasing, the SMs choose a higher *y* to avoid guilt. However, if SMs can correctly use PFI to update *τ''*, then *x* should be correlated with *y*, which it is not. There are alternative explanations to the observed behavior in the lost wallet game, which we discuss below.

A feature of the lost wallet game is that the surplus created by investment is always constant at 20, regardless of *x*. This is different from the well-known investment game (Berg, Dickhaut & McCabe, 1995), in which the amount invested is multiplied by some factor, usually three. Cox, Friedman & Gjerstad (2007), who examine reciprocity in investment games, find a correlation between variables analogous to our *x* and *y*. This suggests that SMs may only be focusing on the amount of surplus generated, and choosing *y* based on how much the FM's action increases their opportunity set (Cox, Friedman & Sadiraj, 2008). As the surplus in the lost wallet game is fixed, this could explain why correlation between *x* and *y* is not observed.[10]

As discussed in the literature review, Ellingsen et al. (2010) suggests that it is not guilt aversion that is being observed in the lost wallet game, rather it is the consensus effect. They propose that *τ''* and *y* are correlated only because SMs are prone to the consensus effect. If the SM believes that the FM would choose the same *y* if the roles were reversed, and if the SM believes the FM bases his *τ'* on what he would have done in the SM's place, then *τ''* would equal *y*, leading to a correlation between them. Such an explanation provides evidence against guilt aversion.

We now present some of the limitations of our work, and this naturally leads to potential avenues for future research. Firstly, we have only tested PFI in the lost wallet game, which, as discussed, may have artefactual issues in its design in regards to surplus generated.

[10] The fixed surplus means that SM behavior in this game could also be explained by distributional preferences such as those outlined in Fehr & Schmidt (1999), and Bolton & Ockenfels (2000). If the SM had, for example, inequality aversion, she would choose *y* close to 10 regardless of *x*, which would result in no correlation between *x* and *y*.

While it was only our intention to investigate the implications for guilt aversion arising from the lost wallet game puzzle, tests in other games related to guilt aversion are required for a more exhaustive conclusion about subject's ability to conduct PFI. Secondly, we did not vary *x*, as in DG, as it was not required to test PFI. An extension where *x* is varied could investigate whether $\tau''$ moves with *x* as it is updated and subsequently affects *y* could provide further evidence for or against guilt aversion. In addition, such design could provide further insight into whether the timing of the belief elicitation influences subject behavior. Finally, while we observed a clear shift in $\tau''$ as elicited by the linear scoring rule, a further extension could be to see if this result is robust to the use of proper scoring rules.

Therefore, based on the results from the current experiment, and in light of the previous literature, we conclude that the puzzling result from DG of no correlation between *x* and *y* in the lost wallet game is not due to a failure of SMs to update their beliefs. It could be an artifact of the game itself, with the surplus fixed at twenty. Alternatively, SMs could simply not be guilt averse, meaning PFI would no longer predict a correlation between *x* and *y*.

**Acknowledgements:** This paper is based on Daniel Woods' Honors thesis written at the University of Canterbury (Woods, 2010). Financial support was provided by the Teaching and Learning Committee, College of Business and Economics, University of Canterbury.

**References**


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

Bellemare, C., Sebald, A., & Strobel, M. (2011). Measuring the willingness to pay to avoid guilt: estimation using equilibrium and stated belief models. *Journal of Applied Economics, 26:3*, 437-453.

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

Bolton, G. E. & Ockenfels, A. 2000. ERC: A Theory of Equity, Reciprocity, and Competition. *American Economic Review, 90:1*, 166-93.

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. (2011). The strategy method versus the direct-response method: a first survey of experimental comparisons. *Experimental Economics, 14,* 375-398.

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 and Partnership. *Econometrica, 74:6*, 1579-1601.

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

Charness, G., Haruvy, E., and Sonsino, D. (2007). Social Distance and Reciprocity: An Internet Experiment. *Journal of Economic Behavior and Organization*, *63:1*, 88-103.

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

Cox, J.C. (2007). A tractable model of reciprocity and fairness. *Games and Economic Behavior, 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 Experimental Lost Wallet Game. *Games and Economic Behavior, 30*, 163-182.

Dufwenberg, M., & Kirchsteiger, G. (2004). A Theory of Sequential Reciprocity. *Games and Economic Behavior, 47*, 268-298.

Dufwenberg, M., Servátka, M. & Vadovič, R. (2015). Honesty and Informal Agreements. Working paper.

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

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

Fehr, E. & Schmidt, K. M. 1999. A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics, 114:3*, 817-68.

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

Gächter, S., & Renner, E. 2010. The effects of (incentivized) belief elicitation in public goods experiments. *Experimental Economics, 13:3*, 364-677.

Geanakoplos, J., Pearce, D., & Stacchetti, E. (1989). Psychological Games and Sequential Rationality. *Games and Economic Behavior*, *1:1*, 60-79.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, *1:1*, 114–125.

Khalmetski, K. (2015). Testing guilt aversion with an exogenous shift in beliefs. Available at SSRN: http://ssrn.com/abstract=2363163 or http://dx.doi.org/10.2139/ssrn.2363163

Khalmetski, K., Ockenfels, A., & Werner, P. (2015). Surprising Gifts: Theory and Laboratory Evidence. *Journal of Economic Theory, 159,* 163-208.

Schlag, K.H., Tremewan, J., & van der Weele, J.J. (2015). A penny for your thoughts: a survey of methods for eliciting beliefs. *Experimental Economics, 18*, 457-490.

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.

Vanberg, C. (2008). Why do people keep their promises? An experimental test of two explanations. *Econometrica*, *76:6,* 1467-1480.

APPENDIX A – Subject Instructions

## INSTRUCTIONS

*(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 $7 and will have two options:

- To keep all of the $7. In this case the paired Group 2 person with whom he/she is paired makes no decision.
- To send all of the $7 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 $7, 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 $7. 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 $7, you should only consider the guesses of the Group 1 people that also sent the $7. If your paired Group 1 person kept the $7, you should only consider the guesses of the Group 1 people that also kept the $7. 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 B – Raw Data**

| FM INFO | |
|---|---|
| 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 |

| FM NO INFO | |
|---|---|
| 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 |

| SM INFO | | |
|---|---|---|
| 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 |

| SM NO INFO | | |
|---|---|---|
| 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 |