Journal of Economic Psychology 56 (2016) 116–125

Contents lists available at ScienceDirect

Journal of Economic Psychology

journal homepage: www.elsevier.com/locate/joep

Testing psychological forward induction and the updating of beliefs in the lost wallet game

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

[a] Department of Economics, Purdue University, United States
[b] MGSM Experimental Economics Laboratory, Macquarie Graduate School of Management, Sydney, Australia

ARTICLE INFO

Article history:
Received 30 October 2015
Received in revised form 2 June 2016
Accepted 10 June 2016
Available online 11 June 2016

JEL classification:
C70
C91

Keywords:
Beliefs
Experiment
Guilt aversion
Lost wallet game
Psychological forward induction
Updating

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 mover's 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.

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, and Stacchetti (1989) and further developed by Battigalli and 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. Bellemare, Sebald, & Strobel, 2011; Charness & Dufwenberg, 2006; Dufwenberg & Gneezy, 2000). 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

* Corresponding author.
E-mail addresses: woodsd42@gmail.com (D. Woods), maros.servatka@mgsm.edu.au (M. Servátka).

[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.

http://dx.doi.org/10.1016/j.joep.2016.06.006
0167-4870/

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, and 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.

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 τ', the SM's belief of τ' as τ'', and an individual's sensitivity to guilt as γ, 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 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 (in this case $\gamma \geqslant 1$), she will send back τ''.

DG's experimental results show that τ'' 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 τ'' conditional on the FM choosing IN should increase, which should in turn lead to a higher y. Given that data from DG shows that τ'' is not correlated with x, it is possible that their subjects fail to psychologically induct forward. They may fail to psychologically induct forward by not updating τ'' to reflect that a FM who chooses IN expects to receive x. If SMs do not correctly update τ'' 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 τ'' 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 τ'' 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 τ''. 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), τ'' 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, and 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, 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

[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 and Kagel (2010). 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).