Games and Economic Behavior 97 (2016) 110–119

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

# Testing guilt aversion with an exogenous shift in beliefs ☆

Kiryl Khalmetski

*University of Cologne, Department of Economics, Albertus-Magnus-Platz, 50923 Cologne, Germany*

ARTICLE INFO

*Article history:*
Received 24 March 2014
Available online 19 April 2016

*JEL classification:*
C91
D82
D83
C72

*Keywords:*
Guilt aversion
Psychological games
Lying

ABSTRACT

We conduct a laboratory experiment to test whether subjects tend to meet the expectations of others (the guilt aversion hypothesis). The specificity of our approach is that second-order beliefs are manipulated exogenously just by changing the parameters of the experimental game. In particular, we consider a simple communication game where the sender is perfectly informed about his material benefit from lying to the receiver. At the same time, the receiver knows only the ex-ante distribution of the sender's material incentives. By changing this distribution between the experimental treatments, we achieve an exogenous variation in the receiver's payoff expectations (and hence in the corresponding sender's second-order beliefs) while keeping the sender's actual material incentives fixed. The results show that the rate of lying is significantly lower when the receiver is supposed to have higher payoff expectations, however only in the case when the monetary incentives for lying are fixed at a moderate level.



## 1. Introduction

A vast economic literature suggests that people care not only about the material consequences of their actions, but also about others' beliefs (Geanakoplos et al., 1989; Dana et al., 2007; Andreoni and Bernheim, 2009). Considerable attention in this field was given to the study of guilt aversion, i.e., an aversion to disappointing others' expectations, which has been shown to have important theoretical implications for strategic behavior (Battigalli and Dufwenberg, 2007, 2009).

However, the experimental evidence on guilt aversion is somewhat mixed, being subject to specific methodological problems. Indeed, guilt aversion implies that individual behavior is affected by second-order beliefs, i.e., beliefs about others' beliefs. At the same time, these beliefs may be endogenous with respect to one's own behavior, i.e., they may simply *follow* behavior because subjects believe that others can predict it well (Vanberg, 2008, p. 1469). On the one hand, this limits the possibility to test guilt aversion by simply eliciting intrinsic second-order beliefs of subjects and then establishing their correlation with prosocial behavior.[1] In particular, if subjects tend to believe that the average behavior is close to their own (the false consensus effect, see Ross et al., 1977), then they might think that the average expectations are also in line

☆ I thank the editor, the advisory editor, two referees, Gary Charness, Martin Dufwenberg, Uri Gneezy, Paul J. Healy, Roman Inderst, Michael Kosfeld, Axel Ockenfels, and seminar and conference participants in Cologne, Kreuzlingen and Maastricht for helpful comments and suggestions. Financial support of the German Research Foundation (DFG) through the Research Unit "Design and Behavior" (FOR 1371) and the ERC (Grant No. 229921) is gratefully acknowledged. Abbreviations: HET – High Expectations Treatment, HIG – High Incentive Game, LET – Low Expectations Treatment, LIG – Low Incentive Game.

*E-mail address:* kiryl.khalmetski@uni-koeln.de.

[1] Such correlation was found to be highly significant in many studies, including Dufwenberg and Gneezy (2000), Guerra and Zizzo (2004), Charness and Dufwenberg (2006), Bacharach et al. (2007), Attanasi et al. (2014).

http://dx.doi.org/10.1016/j.geb.2016.04.003
0899-8256/

with their own actions. This might cause a correlation between second-order beliefs and behavior independently of guilt aversion.[2]

On the other hand, the close link between beliefs and behavior makes it difficult to design an exogenous treatment which does not directly affect behavior, yet significantly shifts second-order beliefs (so that their truly causal effect on behavior can then be traced). Previous studies have proposed different ways to achieve this goal, such as manipulating pre-play communication (Charness and Dufwenberg, 2006; Beck et al., 2013), changing game framing (Dufwenberg et al., 2011), or rescaling of beliefs measurement (Ockenfels and Werner, 2014). However, one might argue that in some of these approaches the manipulated exogenous factor could still cause an additional (potentially interfering) effect on behavior not related to belief-dependent preferences. In particular, Charness and Dufwenberg (2006) used the availability of pre-play communication (specifically, promises) as an instrument to increase first- and second-order beliefs in a trust game. Their treatment with communication was indeed characterized by both higher beliefs and more trustworthy behavior.[3] However, the communication might have a parallel effect on trustees' behavior which is driven by a preference for promise keeping *per se.* This might complicate the estimation of the effect of guilt aversion since the two effects potentially go in the same direction (see, Vanberg, 2008, for a discussion).[4]

The most straightforward approach to exogenously induce second-order beliefs was implemented by Ellingsen et al. (2010) who simply disclosed first-order beliefs of the matched opponents in the trust and dictator games. They found no evidence that second-order beliefs induced in this way affect behavior.[5] Yet, this method has several limitations. First, disclosing individual beliefs of others might signal additional information (besides beliefs *per se*), such as personal traits or social norms, which might also trigger interfering effects (see Khalmetski et al., 2015, for a discussion). Second, from the methodological point of view, since others' beliefs are rarely precisely observable in practice, it is also important to study whether *presumably* different first-order beliefs of the opponent trigger different behavioral responses, which is the approach used in this paper.

In our experiment, second-order beliefs of a subject are manipulated in a direct way just by varying the ex-ante (incomplete) information of his opponent about the game (reflecting its actual stochastic structure), whereas the subject himself remains perfectly informed about the realized game parameters. Hence, the experimental treatment specifically targets second-order beliefs without changing any other aspect of the subject's decision situation (once the parameter realization is fixed), which is the main advantage of the approach. Specifically, we use a simple communication game, where the sender's material incentives to lie are private information of the sender, while the receiver knows only the ex-ante distribution of these incentives. By varying this distribution between the treatments, we exogenously manipulate the receiver's first- and hence the sender's second-order beliefs. At the same time, since the sender is perfectly informed about his actually realized incentives, they can be independently controlled for. Thus, we obtain an exogenous variation in the second-order beliefs for given monetary incentives, which allows us to study the causal effect of these beliefs on behavior (not affected by other effects of the experimental manipulation).[6]

Our results reveal a significant effect of second-order beliefs on the rate of lying in one of the material games. Generally, this provides a clear evidence for guilt aversion, strengthening the previous positive findings (*inter alia*, supporting the expectation-based explanation of the significant treatment effect in Charness and Dufwenberg, 2006). At the same time, we find an additional interaction effect of monetary incentives with guilt aversion. In particular, second-order beliefs do not significantly affect senders' behavior under high monetary incentives for lying. As we elaborate in the discussion section, this suggests that subjects might feel pressure to live up not to *any* expectations of others, but only to those which appear substantiated from their perspective.

The closest paper to ours is that by Ederer and Stremitzer (2015). They also exogenously manipulated first- and second-order beliefs in a variant of the trust game by changing the trustor's ex-ante information about the likely action set of the trustee (instead of changing the ex-ante knowledge about the opponent's incentives as in our paper). Besides, the trustee was given an opportunity to send a free-form message to the trustor at the beginning of the game. They found a significant effect of higher (induced) second-order beliefs on behavior at least for those trustees who have given a promise to the trustor, which is in line with our results. At the same time, there are important differences. First, Ederer and Stremitzer (2015) hypothesized that a decision maker cares about expectations of another player if and only if these expectations are

[2] Empirical evidence for this effect was found in Bellemare et al. (2011), Engelmann and Strobel (2012) and Khalmetski et al. (2015).

[3] The positive effect of the option to make promises on trustworthy behavior was replicated by Ben-Ner et al. (2011), Servátka et al. (2011) and Charness and Dufwenberg (2011), among others.

[4] Vanberg (2008) proposed a method to circumvent this problem by rematching subjects after the communication phase in a dictator game (without informing the recipient whether she has been rematched), to test whether the mere fact that the recipient has been given a promise (from another dictator) affects giving of the lastly matched dictator. Vanberg (2008) found no effect of higher (promise-induced) expectations on giving. However, Kawagoe and Narita (2014) argued that the effect of guilt aversion in this experiment could be affected by the rematching of subjects, which might cause countervailing effects similar to the diffusion of responsibility.

[5] Oppositely, Reuben et al. (2009) and Bellemare et al. (2014) observed a significantly positive correlation between prosocial behavior and disclosed payoff expectations of the opponent. Khalmetski et al. (2015) replicated the setting of Ellingsen et al. (2010) with the strategy method, finding clear evidence for belief-dependent preferences at the within-subject level.

[6] Gneezy (2005) also uses a design where the sender is privately informed about his incentives (while Battigalli et al., 2013, analyze his results in the context of guilt aversion). However, he does not vary the ex-ante distribution of incentives, which drives the treatment effect in our experiment. Costa-Gomes et al. (2014) also use an exogenous shift in payoffs to instrument expectations, yet to study the effect of *first*-order beliefs on behavior.