J Econ Sci Assoc (2017) 3:137–148
https://doi.org/10.1007/s40881-017-0043-0

ORIGINAL PAPER

The limits of guilt

Loukas Balafoutas[1] · **Helena Fornwagner**[1]

Received: 26 June 2017 / Revised: 3 November 2017 / Accepted: 7 November 2017 /
Published online: 16 November 2017
© The Author(s) 2017. This article is an open access publication

Abstract According to the theory of guilt aversion, agents suffer a psychological cost whenever they fall short of other people's expectations. In this paper, we suggest that there may be limits to this kind of motivation. We present evidence from an experimental dictator game showing that behavior is consistent with guilt aversion for relatively low levels of recipient expectations, roughly up to the point where the recipient expects half of the available surplus. Beyond that point the relationship between expectations and transfers becomes negative. Moreover, we examine this relationship at the individual level and establish a typology of subjects depending on how and whether they condition their behavior on recipient expectations.

Keywords Guilt aversion · Experiment · Strategy method · Expectations

JEL Classification C91 · D03

We thank Gary Charness, Martin Dufwenberg, Tore Ellingsen, Magnus Johannesson, Rudolf Kerschbamer, Nikos Nikiforakis, Axel Ockenfels, Ernesto Reuben, Matthias Sutter and Peter Werner for useful comments. We also thank the authors of Ellingsen et al. (2010) for sharing their dataset with us. Financial support from 'eeecon Research Platform' is gratefully acknowledged.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s40881-017-0043-0) contains supplementary material, which is available to authorized users.

✉ Loukas Balafoutas
loukas.balafoutas@uibk.ac.at

[1] Department of Public Finance, University of Innsbruck, Universitaetsstrasse 15, 6020 Innsbruck, Austria

https://doi.org/10.1007/s40881-017-0043-0 Published online by Cambridge University Press

1 Introduction

Human interaction—in families, companies, or clubs—is often influenced by one's perception of other individuals' expectations. It seems that humans have a tendency to feel guilty when they are letting others down, i.e., when their actions do not meet what they believe others expect from them. This human trait has been coined guilt aversion, defined as the emotion that arises when a player 'believes he hurts others relative to what they believe they will get' (Charness and Dufwenberg 2006: 1583).[1] Guilt aversion may influence human behavior in a variety of contexts ranging from marital investments and divorce (Dufwenberg 2002) to corruption in public administration (Balafoutas 2011). In an organizational context, relationships between employers and employees can be shaped by mutual expectations about what constitutes appropriate behavior of either party. In the economic literature, guilt aversion is modeled within the analytical framework of psychological game theory (Geanakoplos et al. 1989; Battigalli and Dufwenberg 2009).

This paper employs a strategy method variant of the dictator game to make two contributions to the literature on guilt aversion and more generally on how social behavior is affected by (perceived) expectations of the involved parties. First, our study explicitly puts forward the idea that the relationship between expectations and behavior is not necessarily monotonic, but instead can have an inverted-U shape, on aggregate as well as at the individual level for some decision makers. We show that dictators display behavior consistent with guilt aversion for relatively low levels of recipient expectations, roughly up to the point where the recipient expects half of the available surplus. Beyond that point, however, the relationship between expectations and transfers becomes negative. This has led us to talk about 'the limits of guilt': the title of this paper aims to convey the intuitive idea that guilt aversion appears to motivate decision makers, but only up to a certain level. When dictators perceive expectations as being too high and therefore illegitimate, they will not attempt to live up to them any longer and they tend to punish recipients who are 'asking too much'.

Second, we establish a typology of subjects based on examination of the relationship between expectations and behavior at the individual level. It would be unreasonable to suggest that every individual's behavior follows the inverted-U shape described above. Accordingly, we classify the 108 dictators who participated in our experiment into six types: selfish types who consistently transfer zero to the recipient; unconditional altruists who give a constant positive amount; positive (or guilt averse) types whose transfers increase with recipient expectations; negative types whose transfers decrease with recipient expectations; hump-shaped types whose transfers increase with expectations up to a certain (individual-specific) level of expectations and decrease beyond that level, meaning that those subjects display the inverted-U shape also at the individual level; and other types who do not fall into any of the five already described categories. We show that positive and negative

[1] Similar definitions can be found in the psychology literature, for instance in Baumeister et al. (1995: 173): 'Feeling guilty [is] associated with···recognizing how a relationship partner's standards and expectations differ from one's own'.

https://doi.org/10.1007/s40881-017-0043-0 Published online by Cambridge University Press

(monotonic) types account for 18% and 20% of subjects, respectively, while a further 20% are classified as hump-shaped.[2]

The above ideas are in line with insights from the existing literature on pro-social behavior, for instance with Charness and Rabin (2005) who argue that how a decision maker responds to the expressed preferences of others depends on how these others have behaved in the past. Similarly, Ghidoni and Ploner (2014) discuss the idea that only legitimate expectations are worth taking into account by a decision maker. The data presented by Andreoni and Rao (2011) reveal that asking for very high amounts can be counter-productive in a setting in which recipients can communicate with dictators. Finally, Regner and Harth (2014) find an inverted-U shaped relationship between second-order beliefs and the amount returned in a trust game.

Experimental evidence on the role of guilt aversion in decision making has been mixed so far. A majority of studies find evidence in favor of guilt aversion in various games (e.g., Dufwenberg and Gneezy 2000; Charness and Dufwenberg 2006; Bacharach et al. 2007; Reuben et al. 2009; Dufwenberg et al. 2011; Attanasi et al. 2013; Beck et al. 2013; Khalmetski et al. 2015; Bellmare et al. 2017a). At the same time, a few papers refute it (Vanberg 2008; Ellingsen et al. 2010—henceforth EJTT; Kawagoe and Narita 2014), show that it is sensitive to context (Balafoutas and Sutter 2016), or find only weak evidence to support it (Charness and Dufwenberg 2010). A crucial methodological issue concerns belief measurement. Guilt aversion means that a decision maker (DM) suffers a psychological cost when she believes she is falling short of the expectations of an affected party (AP). But how should those second-order beliefs be measured experimentally? The approach taken by Charness and Dufwenberg (2006) and others is to elicit the AP's first-order beliefs and then ask the DM to estimate them. This seems like a natural way to elicit second-order beliefs, but it is vulnerable to a false consensus effect. EJTT overcome this problem by eliciting first-order beliefs and then directly transmitting them to the DM. However, it is possible that (some of) the affected parties report beliefs in a strategic manner, for instance, if they believe that guilt averse decision makers would then make higher transfers. Moreover, dictators know that there are undisclosed design features, which may raise suspicion and result in loss of experimental control. In this paper we follow the approach of EJTT, acknowledging, however, that both methods have their strengths and weaknesses.

Our method is based on a dictator game, in which we ask dictators to report a transfer for each possible first-order belief of the recipient that she is matched with. This technique is akin to a strategy method, since it conditions choices on a co-player's beliefs. Its main advantage is that it allows us to exclude the possibility of a false consensus effect and at the same time to elicit a profile of transfers from the dictator. This method has previously been used in Khalmetski et al. (2015), henceforth KOW, who find that the relationship between dictator giving and

[2] In a recent experiment conducted in parallel to our work, Bellemare et al. (2017a) examine sensitivity to guilt aversion at the individual level in a dictator game, finding substantial heterogeneity and dependence on the level of stakes involved. Attanasi et al. (2013) provide a categorization of second-movers in a trust game based on the relationship between perceived beliefs and back-transfers and classify around 55% of subjects as guilt averse.

https://doi.org/10.1007/s40881-017-0043-0 Published online by Cambridge University Press

recipient expectations is positive for some dictators and negative for others. What differs, however, is the interpretation of the data. In the model of KOW dictators may have a disutility from creating negative surprises, which leads to a positive relationship between expectations and transfers in line with guilt aversion. But dictators also draw utility from creating positive surprises: the lesser recipients expect, the greater the positive surprise dictators can create and hence the more they are inclined to give. This latter motive can lead to a negative relationship between expectations and transfers, in line with our results. While we consider this a plausible and interesting story, we note that it is inconsistent with a hump-shaped relationship at the individual level and hence cannot explain the behavior of a substantial fraction of dictators in our sample. Moreover, we go one step further and analyze the relationship between transfers and beliefs at the individual level with the aim of classifying dictators into different types depending on their underlying motivation. Hence, we view our results as complementary to KOW.[3]

2 Experimental design and procedures

Subjects were randomly assigned to one of two types, dictators or recipients, located in two different rooms. Dictators received an endowment of €16, while recipients received no endowment, but were paid a show-up fee of €5.[4] Each dictator was then asked to decide how much of their endowment to transfer to the recipient that she had been randomly matched with. Possible transfers included every amount between €0 and 16€ (in €1 steps), including €0 and €16. Recipients were not able to act at any time during the experiment but were asked about their expectation of the average transfer that dictators would give to recipients within the session. These first-order beliefs were incentivized: the recipient whose expectation was closest to the actual average transfer in the session received €12 in addition to his realized transfer.[5] If there was more than one correct estimate, the winner was chosen by chance. At this point we acknowledge that introducing a payment for correct estimates could lead to a bias if subjects hedge their experimental income using their stated estimate (Blanco et al. 2010). However, as EJTT note, subjects state their belief about the average realized transfer and the stakes are small. Therefore, the probability of hedging incomes is mitigated. Further, hedging would only become a problem if the dictators believe that recipients hedge instead of stating their true belief.

[3] Hauge (2016) also employs a strategy method variant of the EJTT experiment in which dictators choose their transfers conditional on three possible belief levels (0, 50% of the surplus, or a level in-between). She finds a positive relationship between transfers and beliefs, which is fully consistent with our findings up to a belief of 8 (50% of the surplus). However, she does not consider higher levels of beliefs, for which we find a negative relationship.

[4] The fact that recipients received a show-up fee was made necessary due to the rules of our lab. However, this was not mentioned to dictators for two reasons: First, to ensure that transfers as well as beliefs would not be affected by the show-up fee asymmetry; and second, to ensure a greater comparability of our findings with existing papers such as EJTT and KOW.

[5] This was the average of all realized transfers, which were determined based on each dictator's donation profile and his or her matched receiver's stated first-order belief.

https://doi.org/10.1007/s40881-017-0043-0 Published online by Cambridge University Press

Following KOW, we employed a design akin to the strategy method for dictator decisions. In particular, dictators had to fill out a table where they stated their transfer for every possible expectation (i.e., for each elicited first-order belief) of their recipient (varying from €0 to €16). This methodology allows us to elicit a full profile of transfers from each dictator, for each belief level. Dictators were informed after filling out the table what the estimate of their matched recipient was, and depending on this estimate the relevant transfer was actually implemented.[6]

Subjects of both types were subsequently asked to fill out a questionnaire with socio-demographic questions and completed a ten-question version of the Big-5 personality questionnaire (Gosling et al. 2003). Payments were made anonymously in cash and averaged €12.50 for dictators and €9.06 for recipients. All sessions were conducted at the EconLab of the University of Innsbruck using paper and pen and lasted for around 40 min. We recruited 216 students of different academic backgrounds using H-Root (Bock et al. 2014). We ran five sessions in total, four of them with 44 subjects and one with 40 subjects. This means that we have data for 108 dictators and 108 recipients.

3 Results

3.1 Aggregate analysis

Overall, the mean conditional transfer in our experiment is €3.23, which amounts to 20% of the total available surplus of €16. This is very close to the averages reported in EJTT ($3.60; 24% of the endowment) and KOW (€3.25; 23% of the endowment). To better illustrate the comparability of our findings with those obtained by means of the direct response method in EJTT, Table 1 compares average transfers for various ranges of beliefs in the two studies. The table reveals that average transfers are very similar, both in terms of levels and in terms of the general pattern (first increasing, and then decreasing for high levels of beliefs).

Figure 1 plots the mean transfer conditional on each level of beliefs. This figure reveals that the relationship between beliefs and dictator giving has an inverted-U shape, with transfers roughly increasing up to a belief of eight (Spearman's $\rho = 0.13$, $p < 0.01$) and then decreasing for the remaining range of beliefs ($\rho = -0.06$, $p = 0.06$). It follows that, from the point of view of a recipient, the optimal strategy would be to report an intermediate belief: transfers are highest when beliefs are exactly at the equal split of eight ($t_8 = 3.75$) and lowest when the recipient expects a transfer of one ($t_{15} = 2.50$). Table A1 in the online

[6] While the strategy method might be prone to demand effects, we note that our results are—to the extent comparable—fully consistent with EJTT where the direct response method is used. Furthermore, numerous studies like Brandts and Charness (2011), Fischbacher et al. (2012), and KOW find no evidence that the two methods yield qualitatively different results. A recent paper by Bellemare et al. (2017b) compares three different methods for testing guilt aversion in a dictator game. The findings of that paper reveal that the strategy method yields very similar results to those obtained when second-order beliefs of dictators are elicited, and that the method based on disclosing recipients' first-order beliefs used by EJTT produces different results compared to the other two methods.

https://doi.org/10.1007/s40881-017-0043-0 Published online by Cambridge University Press

Table 1 Comparison of transfers with Ellingsen et al. (2010)

Belief range, as share of the total surplus	Average transfer, as share of the total surplus	
	Ellingsen et al. (2010)	This study
0	0.167	0.161
(0, 0.20)	0.214	0.181
[0.20, 0.40)	0.281	0.225
[0.40, 1]	0.215	0.205

In our paper the number of observations is equal in all categories thanks to the use of the strategy method, $N = 108$. In Ellingsen et al. (2010) the number of observations varies depending on how many reported beliefs fall in each category: for the four categories {0; (0, 0.20); [0.20, 0.40); [0.40, 1]} N equals 4, 21, 35, 24, respectively

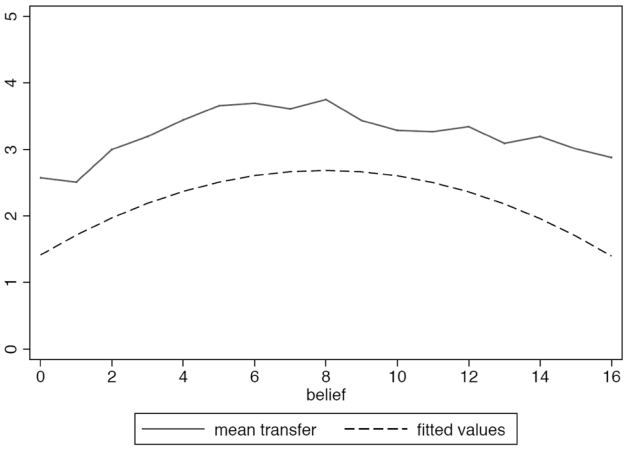


Fig. 1 Mean transfer, by belief

appendix shows the exact mean transfers by belief level, while for completeness in the appendix (Figure B1) we also report the first order beliefs given by recipients.

The inverted-U shape in the relationship between dictator giving and recipient expectations may help explain why a number of papers fail to detect a significant relationship between giving and beliefs, since the increasing and the decreasing part of this relationship are likely to cancel each other out. As a matter of fact, in our experiment we also find no significant correlation between giving and beliefs over the entire range of beliefs (Spearman's $\rho = -0.01$, $p = 0.79$). Hence, had we only tested for a positive relationship, we would have failed to find one and would have concluded that guilt aversion does not drive dictators' giving decisions. Hence, the aggregate analysis of our data points towards a potential explanation for the conflicting findings on the relationship between expectations and behavior in the literature.

https://doi.org/10.1007/s40881-017-0043-0 Published online by Cambridge University Press

Table 2 Regression results

Dependent variable: dictator transfer	(1)	(2)
Belief	0.319***	0.329***
	(0.049)	(0.052)
Belief^2	− 0.020***	− 0.021***
	(0.003)	(0.003)
Female dictator		− 0.536
		(1.041)
Age		− 0.203
		(0.177)
Extraversion		0.028
		(0.371)
Agreeableness		0.713
		(0.497)
Neuroticism		0.342
		(0.363)
Conscientiousness		− 0.207
		(0.404)
Openness		0.374
		(0.449)
Constant	1.414***	0.391
	(0.447)	(6.079)
N	1836	1751

Tobit regressions with dictator random effects

Dependent variable left-censored at 0

Standard errors are clustered by subject and shown in parentheses

***Denotes significance at the 1% level. As five subjects did not fill out the Big Five Questionnaire, the number of observations is lower in specification (2)

Table 2 shows the results of Tobit regressions with individual transfers as the dependent variable. The right-hand side variables are the level of the recipient's belief and its square, to control for quadratic effects indicative of an inverted-U shape, as well as age, gender and Big 5 personality traits in specification (2). In both specifications we obtain the predicted positive coefficient for the linear term and negative coefficient for the quadratic term, both significant at the 1% level. The fitted values from specification (1) are plotted in Fig. 1: the global maximum is estimated at a belief level of 7.72, which is in line with the actual data. In (2) we include our controls without finding any notable changes in our coefficients of interest. As with Fig. 1, the main purpose of the regressions is to illustrate the fact that the relationship between beliefs and donations is not a linear one. At the same time, this kind of analysis does not take into account the possibility that the estimated coefficients are capturing several (potentially opposite) effects of beliefs for different dictators, and therefore, it masks the individual heterogeneity that gives rise to these effects and to this aggregate pattern. For this reason, in the following section we examine the relationship between beliefs and donations at the individual level.

3.2 Individual-level analysis and typology of subjects

In this part, we turn to the analysis of the donation profiles of dictators at the individual level. For this purpose, we have plotted the relationship between beliefs

https://doi.org/10.1007/s40881-017-0043-0 Published online by Cambridge University Press

and transfers for each dictator and include them in Figure B2 in the Appendix. Based on the observed patterns of behavior, we have classified dictators into one of six distinct behavioral types:

1. Selfish types whose transfers are constant at zero and independent of the recipient's beliefs, with a maximum of one deviation to a positive transfer over the 17 decisions.
2. Unconditional altruists who transfer a constant positive amount independent of beliefs.
3. Positive (guilt averse) types whose transfers are positively correlated to recipients' expectations. Following the seminal work by Fischbacher et al. (2001) who classify subjects into four behavioral types based on their strategy profile in a public goods game, we rely on the Spearman rank correlation coefficients and classify a subject as guilt averse if the correlation between transfers and beliefs is positive and significant at least at the 5% level.[7]
4. Negative types whose transfers are negatively correlated with recipients' expectations (with Spearman's ρ significant at 5%).
5. Hump-shaped types whose transfers are positively correlated with expectations up to a certain threshold, or switching point called S_i, and negatively correlated with expectations beyond S_i (with Spearman's ρ significant at 5% for both). To identify these subjects we looked for possible S_i's which would satisfy this condition for each subject, and classified a subject as hump-shaped if such a S_i existed.
6. Other types who do not fall into any of the categories (1)–(5) above.

Hence, two of the above types (selfish subjects and unconditional altruists) do not condition their transfers on the expectations of the recipient, while the opposite is true for types (3)–(5). Those types condition their transfers on expectations in a systematic way, either positively, negatively, or both. Table 3 shows the distribution of the six types within the entire population of dictators. The first thing to note is that 20.4% of subjects do not condition their transfers on the expectations of the recipient. Of those, 13.9% are selfish and 6.5% are unconditional altruists.[8] On the contrary, 58.3% of all subjects conditioned their transfers on expectations in a systematic way. Among those subjects we find a slightly smaller number of guilt-averse subjects (with a positive slope in their profile of transfers) than of subjects with a negative slope, with the two types accounting for 17.6 and 20.4% of the sample, respectively. A further 20.4% of subjects can be classified as hump-shaped, i.e., as displaying a positive relationship up to a switching point S_i and a negative one beyond that point. Of course, every one of those dictators may differ with

[7] Fischbacher et al. (2001) use the 1% significance level as a requirement for their classification. In the Appendix (Table A2) we present a version of Table 3 in which we use $p < 0.01$ for classification. Naturally, this more stringent criterion increases the proportion of subjects who cannot be allocated to one of the five main categories and fall into the category of 'other types'. This affects the classification of nine subjects in total.

[8] Of the 15 subjects that we classify as selfish, three chose a positive transfer (usually €1) in one of their 17 decisions. Of the seven subjects that we classify as unconditional altruists, two always chose a transfer of 8 (the equal split) or 1, and the transfer levels of 2, 4 and 6 were each chosen by one subject.

https://doi.org/10.1007/s40881-017-0043-0 Published online by Cambridge University Press

Table 3 Distribution of types

Person's type	Freq.	Percent
Selfish	15	13.89
Unconditional altruist	7	6.48
Positive	19	17.59
Negative	22	20.37
Hump-shaped	22	20.37
Other	23	21.30
Total	108	100

respect to their switching point S_i. In particular, among the 22 subjects in this category, the distribution of the identified levels for S_i is as follows: the mode lies at the equal split of $S_i = 8$ for eight subjects, while two subjects have their switching point at $S_i = 7$ and one subject at $S_i = 9$, meaning that 50% of subjects who belong to that type have their switching point at or around the equal split. Two further subjects switch already at $S_i = 3$, one subject switches at $S_i = 4$, three switch $S_i = 5$, and five subjects switch at $S_i = 6$, respectively.[9]

While the frequencies of the various types listed in Table 3 nicely illustrate the heterogeneity of individual donation patterns and motivations, we note that the absolute levels of these frequencies should be interpreted with caution since they are likely to depend on the subject pool or the recruitment procedures in the particular experiment. A recent literature has examined whether there is a selection bias into economic experiments, which would imply that participants are not representative of the student population from which they are drawn: while Cleave et al. (2013) and Falk et al. (2013) find no evidence of a selection bias with respect to pro-social inclination, Slonim et al. (2013) show that lab participants are unrepresentative of the student population along a number of relevant characteristics. Moreover, Eckel and Grossman (2000) report that the recruitment method in lab experiments has a substantial impact on altruistic behavior. A further critical issue is the possibility that student samples are unrepresentative of the general population (see, e.g., Anderson et al. 2013; Falk et al. 2013). Hence, it is important to keep in mind that the precise distribution of types in the general population is likely to deviate from the values shown in Table 3 due to a number of possible factors relating to subject pools, recruitment methods, or other aspects of the experimental design.

4 Concluding remarks

The goal of this paper has been to contribute to the literature on guilt aversion by suggesting that the relationship between a decision maker's behavior and an affected party's perceived expectations need not be monotonic. We have used a

[9] It must be noted that our classification method may be underestimating the true proportion of hump-shaped types to some extent. This can be the case for subjects who have their switching point quite early (or quite late), such that no significant positive (negative) relationship is documented up to (beyond) that point. Such subjects would be classified as negative or positive types, instead of hump-shaped.

https://doi.org/10.1007/s40881-017-0043-0 Published online by Cambridge University Press

strategy method variant of the dictator game and shown that mean transfers across dictators increase with recipient expectations up to a certain threshold but decrease beyond that threshold. Furthermore, we have been able to classify dictators into a number of different types depending on the sign of the slope of this relationship in their elicited donation profile and have found that around six out of ten dictators condition their giving on recipient expectations, either acting in line with guilt aversion, reducing their transfers as expectations increase, or both.

We believe that, by suggesting that there is a threshold beyond which guilt aversion no longer applies and higher perceived expectations lead to less kind behavior on the part of the decision makers, we are offering an important insight which may help reconcile some of the controversy in the literature on guilt aversion. Nevertheless, certain limitations need to be pointed out. For one, we cannot be sure that the mechanism driving the negative part in the relationship between giving and beliefs is due to a motive for punishing recipient expectations that are too high and illegitimate as seen from the perspective of the dictator. We readily acknowledge that more evidence is needed to corroborate this phenomenon. For instance, one obvious step would be to look for evidence of a role for (un)acceptable expectations in different contexts, such as trust games.[10] In any case, we consider our data pattern a very interesting empirical regularity that deserves to be further investigated in future studies.

Acknowledgements Open access funding provided by University of Innsbruck and Medical University of Innsbruck.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Anderson, J., Burks, S., Carpenter, J., Götte, L., Maurer, K., & Nosenzo, D. (2013). Self-selection and variations in the laboratory measurement of other-regarding preferences across subject pools: evidence from one college student and two adult samples. *Experimental Economics, 16,* 170–189.

Andreoni, J., & Rao, J. (2011). The power of asking: how communication affects selfishness, empathy, and altruism. *Journal of Public Economics, 95,* 513–520.

Attanasi, G., Battigalli, P., Nagel, R., (2013). *Disclosure of belief-dependent preferences in a trust game*. IGIER Working Paper 206, Bocconi University.

[10] To give one concrete example, a motive for punishing illegitimate expectations is fully consistent with some of the data patterns presented in Charness and Dufwenberg (2006). Comparing treatments (5, 5) and (7, 7) based on game Γ_1 of that paper we see that player B is less trustworthy in treatment (7, 7) when the outside options are higher. The authors say that '*perhaps this is…unexpected*' (p. 1588), but we argue that it is reasonable if we consider the idea of legitimate expectations. By playing 'In' in (5,5), player A is in effect expecting B to give up 4 so that A can gain 5 (in expected terms). In (7, 7) A is in effect asking B to give up 4 so that A can gain only 3, so the lower trustworthiness of player B in this case may be because B thinks that A is asking too much.

https://doi.org/10.1007/s40881-017-0043-0 Published online by Cambridge University Press

Bacharach, M., Guerra, G., & Zizzo, D. (2007). The self-fulfilling property of trust: an experimental study. *Theory and Decision, 63,* 349–388.
Balafoutas, L. (2011). Public beliefs and corruption in a repeated psychological game. *Journal of Economic Behavior & Organization, 78,* 51–59.
Balafoutas, L., Sutter, M., (2016). On the nature of guilt aversion: Evidence from a new methodology in the dictator game. *Journal of Behavioral and Experimental Finance*, forthcoming.
Battigalli, P., & Dufwenberg, M. (2007). Guilt in games. *American Economic Review, Papers and Proceedings, 97,* 170–176.
Battigalli, P., & Dufwenberg, M. (2009). Dynamic psychological games. *Journal of Economic Theory, 114,* 1–35.
Baumeister, R., Stillwell, A., & Heatherton, T. (1995). Personal narratives about guilt: role in action control and interpersonal relationships. *Basic and Applied Social Psychology, 17,* 173–198.
Bellemare, C., Sebald, A., Suetens, S., (2017a). Heterogeneous guilt sensitivities and incentive effects. *Experimental Economics*, forthcoming.
Bellemare, C., Sebald, A., & Suetens, S. (2017b). A note on testing guilt aversion. *Games and Economic Behavior, 102,* 233–239.
Blanco, M., Engelmann, D., Koch, A. K., & Normann, H. T. (2010). Belief elicitation in experiments: is there a hedging problem? *Experimental Economics, 13,* 412–438.
Bock, O., Baetge, I., & Nicklisch, A. (2014). Hroot—hamburg registration and organization online tool. *European Economic Review, 71,* 117–120.
Brandts, J., & Charness, G. (2011). The strategy versus the direct-response method: a first survey of experimental comparisons. *Experimental Economics, 14,* 375–398.
Charness, G., & Dufwenberg, M. (2006). Promises and partnership. *Econometrica, 74,* 1579–1601.
Charness, G., & Dufwenberg, M. (2010). Bare promises: an experiment. *Economics Letters, 107,* 281–283.
Charness, G., & Rabin, M. (2005). Expressed preferences and behavior in experimental games. *Games and Economic Behavior, 53,* 151–169.
Cleave, B., Nikiforakis, N., & Slonim, R. (2013). Is there selection bias in laboratory experiments? The case of social and risk preferences. *Experimental Economics, 16,* 372–382.
Dufwenberg, M. (2002). Marital investments, time consistency and emotions. *Journal of Economic Behavior & Organization, 48,* 57–69.
Dufwenberg, M., Gächter, S., & Hennig-Schmidt, H. (2011). The framing of games and the psychology of play. *Games and Economic Behavior, 73,* 459–478.
Dufwenberg, M., & Gneezy, U. (2000). Measuring beliefs in an experimental lost wallet game. *Games and Economic Behavior, 30,* 163–182.
Eckel, C., & Grossman, P. (2000). Volunteers and pseudo-volunteers: the effect of recruitment method in dictator experiments. *Experimental Economics, 3,* 107–120.
Ellingsen, T., Johannesson, M., Tjotta, S., & Torsvik, G. (2010). Testing guilt aversion. *Games and Economic Behavior, 68,* 95–107.
Falk, A., Meier, S., & Zehnder, C. (2013). Do lab experiments misrepresent social preferences? The case of self-selected student samples. *Journal of the European Economic Association, 11,* 839–852.
Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters, 71,* 397–404.
Fischbacher, U., Gächter, S., & Quercia, S. (2012). The behavioral validity of the strategy method in public good experiments. *Journal of Economic Psychology, 33,* 897–913.
Geanakoplos, J., Pearce, D., & Stacchetti, E. (1989). Psychological games and sequential rationality. *Games and Economic Behavior, 1,* 60–79.
Ghidoni, R., Ploner, M., 2014. When do the expectations of others matter? An experiment on distributional justice and guilt aversion. CEEL Working Paper 3–14.
Gosling, S., Rentfrow, P., & Swann, W., Jr. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37,* 504–528.
Hauge, K. (2016). Generosity and guilt: the role of beliefs and moral standards of others. *Journal of Economic Psychology, 54,* 35–43.
Kawagoe, T., & Narita, Y. (2014). Guilt aversion revisited: an experimental test of a new model. *Journal of Economic Behavior & Organization, 102,* 1–9.
Khalmetski, K., Ockenfels, A., & Werner, P. (2015). Surprising gifts: theory and laboratory evidence. *Journal of Economic Theory, 159,* 163–208.

https://doi.org/10.1007/s40881-017-0043-0 Published online by Cambridge University Press

Regner, T., Harth, N., 2014. Testing belief-dependent models. Max-Planck Institute of Economics Jena Working Paper.
Reuben, E., Sapienza, P., & Zingales, L. (2009). Is mistrust self-fulfilling? *Economics Letters, 104,* 89–91.
Slonim, R., Wang, C., Garbarino, E., & Merrett, D. (2013). Opting-in: participation bias in economic experiments. *Journal of Economic Behavior & Organization, 90,* 43–70.
Vanberg, C. (2008). Why do people keep their promises? An experimental test of two explanations. *Econometrica, 76,* 1467–1480.

https://doi.org/10.1007/s40881-017-0043-0 Published online by Cambridge University Press