No. 90

UNIVERSITY OF COLOGNE

WORKING PAPER SERIES IN ECONOMICS

SHARING GUILT: HOW BETTER ACCESS TO INFORMATION MAY BACKFIRE

ROMAN INDERST
KIRYL KHALMETSKI
AXEL OCKENFELS

Department of Economics
University of Cologne
Albertus-Magnus-Platz
D-50923 Köln
Germany

http://www.wiso.uni-koeln.de

Sharing Guilt:

How Better Access to Information May Backfire

Roman Inderst[1]

Goethe University Frankfurt, Theodor-W.-Adorno-Platz 1, D-60323 Frankfurt am Main, Germany

Kiryl Khalmetski

University of Cologne, Albertus-Magnus-Platz, D-50923 Cologne, Germany

Axel Ockenfels

University of Cologne, Albertus-Magnus-Platz, D-50923 Cologne, Germany

2 May 2017

Abstract

We study strategic communication between a customer and an advisor who is privately informed about the best suitable choice for the customer, but whose preferences are misaligned with the customer's preferences. The advisor sends a message to the customer who, in turn, can secure herself from bad advice by acquiring costly information on her own. We find that making the customer's information acquisition less costly, e.g., through consumer protection regulation or digital information aggregation and dissemination, leads to *less* prosocial behavior of the advisor. This can be explained by a model of shared guilt, which predicts a shift in causal attribution of guilt from the advisor to the customer if the latter could have avoided her *ex post* disappointment.

Keywords: shared guilt, trust, guilt aversion, responsibility diffusion, advice

JEL classification: C91, D82, D83

[1] We thank Martin Dufwenberg, Nicholas Epley and seminar and conference participants in Cologne, Gothenburg and Kreuzlingen for helpful comments. 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.

1. Introduction

Informational transparency is often viewed as a fundamental prerequisite for efficient trading and a smooth functioning of markets and institutions (Akerlof 1970, Stiglitz and Weiss 1981, Kaufmann and Bellver 2005). For instance, much effort in consumer protection regulation has been put into making it easier for consumers to directly assess and compare the various products and options that are available to them, rather than having to rely on the advice of sellers or brokers who often have conflicting preferences. Examples of this approach abound and include the standardization of product information, the presentation of information in transparent language, or the use of labels (Golan et al. 2001, Kozup and Hogarth 2008). Similarly, recent advances in computer technology have helped to substantially reduce consumers' uncertainty about product quality, sellers' trustworthiness and competitors' price offers in digital markets and e-commerce (Bakos 1997, Goldmanis et al. 2010, Bolton et al. 2013). Such regulation and digitalization of economic interaction are supposed to help customers make better decisions.

In this paper, however, we show that, if social preferences come into the picture, informational transparency might backfire. To the extent it is made easier for customers to protect themselves against bad decisions, professional advisors may feel less socially responsible to help out. We provide controlled laboratory evidence that a policy that facilitates information transparency may crowd out social responsibility of advisors, and propose a model that describes the underlying behavioral mechanism.[2]

In particular, we consider a communication game where an uninformed customer can refer to an informed advisor before taking the product choice. If the customer refers to the advisor, the latter has strict monetary incentives to lie to the customer. At the same time, the customer has an option to acquire information on her own at a fixed cost. We find that the immediate benefits from easing access to product information, by lowering the costs to customers of getting fully informed by themselves, may be overcompensated. The reason is that when customers face lower costs of obtaining information but still rely on the advisor, they are more

[2] This countervailing effect of an informational transparency policy is difficult to identify in a naturally occurring field setting. The only exception we are aware of is Ahmad et al. (2006), who show that physicians may become less helpful to a patient once they learn that he or she is using the Internet to access health information from different sources. While we focus on cases where customers can make good use of an informational transparency policy, such policy has also been criticized in instances where consumers are expected to make little use of this additional information, e.g., due to information overload (Lacko and Pappalardo 2007). Other contributions focus on greater disclosure of conflicts of interest, arising for instance from commissions, and on how this affects firms' strategic behavior (e.g., Inderst and Ottaviani 2012, Anagol and Kim 2012, Duarte and Hastings 2012).

likely to receive unsuitable advice than when obtaining outside information involves larger costs. That is, if customers forgo the opportunity to inform themselves at relatively low costs, advisors are less willing to act in the consumers' interest, and rather lie, even though they rightly anticipate that consumers who self-select into seeking advice put more trust in them.

On a conceptual level, we show that the observation can be accommodated by a natural extension of the concept of guilt in psychological game theory. Psychological game theory (Geanakoplos et al. 1989, Battigalli and Dufwenberg 2009) postulates that individual preferences accommodate not only monetary payoffs, but also beliefs (including the beliefs about others' beliefs, and so on). This way, it can capture that behavior depends on foregone outside options. However, applying the canonical model of guilt aversion (Battigalli and Dufwenberg 2007, 2009) to our context would suggest that a better foregone outside option of the customer would unambiguously lead to *less* lying – the opposite of our finding. The reason is that the customer's choice to forego this option reveals her higher expectation of the advisor's trustworthiness, who in turn would be less inclined to lie because this would yield a stronger customer's disappointment and thus generate more guilt according to the model. We extend this model to allow for a shift of causal attribution of guilt from the advisor to the customer if the latter could have avoided the loss. The underlying idea is that guilt is *shared* by the players according to the extent each player can be blamed for the outcome. We show how shared guilt allows for a negative relation between the customer's cost to choose the (foregone) outside option and the advisor's tendency to lie. There are two conflicting forces at work when information could have been obtained at a smaller cost, yet the customer nevertheless relies on the advisor: (*i*) more revealed trust in the advisor and thus more guilt from lying, and (ii) a shift of the attribution of guilt to the customer. Our experimental setting allows us to disentangle these two conflicting forces. In a second experiment, we can also rule out other signaling and self-selection effects. The results from both of our experiments are in line with our model of shared guilt.

Our hypothesis is related to what Charness and Rabin (2002) called the "complicity effect", which posits that the mere fact that a trustor is an active player might diffuse the trustee's responsibility for the final outcome. Indeed, in their experiment they observed no evidence that a trusting choice of the trustor leads to more prosocial behavior of the trustee.[3] These results are consistent with our findings and model, yet an important difference of our approach is that we observe the effect of responsibility diffusion by varying *only* the outside option of the

[3] See also Brandts et al. (2015) who observed no evidence for reciprocating trust in a similar experiment.

customer (i.e., the trustor) while keeping her actual choice (to trust) fixed, thus controlling for an important element of positive reciprocity.[4]

More generally, we relate to earlier results of Charness (2000) who asserted and experimentally showed that individuals behave more prosocially in situations where they have full responsibility for the outcome, and correspondingly less so if this responsibility can be shared with another party.[5] This effect was further confirmed in studies of strategic delegation which showed that responsibility can be diffused among several players (jointly) causing a harm to another player through delegation of decisions (Fershtman and Gneezy 2001, Hamman et al. 2010, Coffman 2011, Bartling and Fischbacher 2012, Oexl and Grossman 2013).[6] Unlike the approaches taken in this literature, we involve the negatively affected player into the sharing of responsibility.[7]

Our model and experiments reconcile seemingly inconsistent results in the experimental literature on belief-dependent preferences. On the one hand, various studies (e.g., Charness and Dufwenberg 2006, Reuben et al. 2009, Khalmetski 2016) showed that higher second-order beliefs *all else equal* lead to more trustworthiness. This is consistent with both the standard model of guilt aversion and our model of shared guilt, as well as with our experimental results. On the other hand, multiple experiments observed that more trusting *behavior*, i.e., behavior signaling higher expectations, does not lead to more trustworthiness (Dufwenberg and Gneezy 2000, Servátka & Vadovič 2009, Cox et al. 2010, Beck et al. 2013, Woods and Servátka 2016). This is inconsistent with what is suggested by the standard model of guilt aversion, yet consistent with our extended model and also our experimental results.

Our work also contributes to the experimental literature on the determinants of subjects' tendency to lie, starting with Gneezy (2005). One of the key findings in this literature is that lying is not only affected by the monetary consequences for the liar, but also by those for the

---

[4] The experiment of Dufwenberg and Gneezy (2000) was also based on varying the outside option of a trustor in a variant of the trust game (the Lost Wallet Game). At the same time, the payoff structure in their experiment implied that the trusting choice of the trustor was explicitly kind to the trustee (as otherwise he got 0 payoff). This allowed for a larger scope for positive reciprocity than in our experiment, where the payoff structure was adjusted to disentangle the effect of shared guilt.

[5] Charness (2000) termed this as "the responsibility-alleviation effect". Charness et al. (2012) further showed that subjects might even *strategically* delegate the full responsibility over outcome to another party in order to trigger a prosocial response (i.e., by strategically reducing the alleviation of responsibility).

[6] Garofalo and Rott (forthcoming) showed that decision makers may sometimes attempt to shift blame by the delegation of *communicating* an unfair allocation (unlike delegation of the decision rights as in the aforementioned literature). Interestingly, such attempts are generally not successful as they do not lead to less punishment of the decision maker from the negatively affected players, and can even backfire.

[7] Bartling and Fischbacher (2012) also propose a model of diffusion of responsibility between active players whose decisions affect another party based on a belief-based measure of anticipated responsibility, which bears similarities to the way we model shared guilt preferences. Yet, Bartling and Fischbacher (2012) do not consider a dependence of responsibility attribution on second-order beliefs, which is at the core of our model.

receiver of the message (Gneezy 2005, Erat and Gneezy 2012). Our experiments add that not only materialized consequences but also unrealized consequences (outside options) of the receiver matter in systematic and probably unexpected ways for the propensity to lie.

Finally, returning to our motivating discussion, our results show that increased informational transparency can make more vulnerable those customers, who prefer to continue to trust in others' advice. As far as advisors' conduct is not completely constrained by rules and liability concerns, their advice to customers may become more biased, because of a shift of blame which crowds out advisors' social concerns for customers. While a policy implemented to ease access to information may thus have the intended consequences for those customers who in fact sidestep advice, it may backfire for those customers who still rely on it.[8]

Section 2 presents our theoretical setting. Section 3 describes the baseline Experiment 1. Section 4 provides the results of a robustness check, Experiment 2, and Section 5 concludes.

2. Modeling shared guilt

2.1. Monetary payoffs in the Sharing Guilt Game

Suppose there are two players, an advisor (he) and a customer (she), playing the Sharing Guilt Game shown in Fig. 1. In the first stage, the customer chooses between *In* and *Out*. In case of *Out* the game ends and the customer gets her outside option $\gamma \in (0,1)$, while the advisor gets a low payoff π_a^L. In case of *In*, the game proceeds to the next stage where the advisor chooses between two actions: *Truth* or *Lie*. In the final stage, the customer chooses between *Follow* or *Not Follow*, not knowing whether the advisor lied or not. If the advisor tells the truth, the customer gets a payoff normalized to 1 if she follows the advisor, and a payoff of 0 if she does not follow him. If the advisor lies, the customer gets 0 if she follows the advisor. If she does not follow him in this case, she gets 1 with a small probability ϵ, and 0 with the remaining probability. One interpretation is that in this case the customer relies on her own limited knowledge in taking some final decision, or she just makes some random decision and hopes for good luck, so expected payoffs are small. The advisor earns a high monetary payoff $\pi_a^H > \pi_a^L$ if and only if the customer gets 0. With these monetary payoffs and sufficiently small ϵ, in subgame perfect equilibrium, the customer plays *Out*.[9]

---

[8] Another backfiring mechanism through less suitable advice has been recognized by Cain et al. (2005, 2011), yet in a very different context: They show that disclosing a conflict of interest between advisor and advisee can induce less trustworthy behavior of advisors.

[9] Indeed, the customer can earn sufficiently high payoff after *In* only if she subsequently plays *Follow* with a non-negligible probability. However, then the advisor will always lie which leads to zero payoff of the customer.

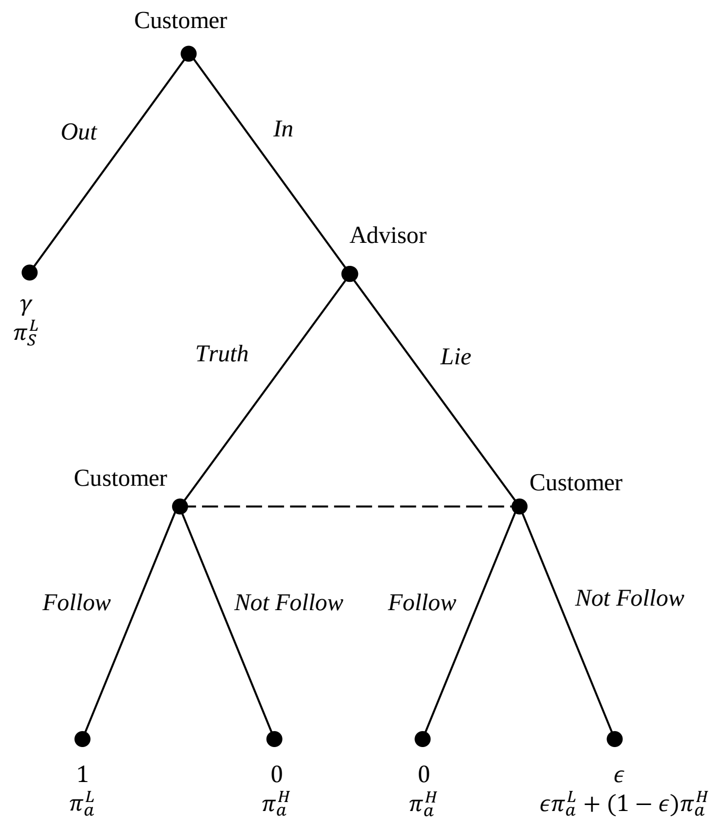

Fig. 1. Monetary Payoffs in the Sharing Guilt Game.

The game reflects situations in which players have opposing interests, and thus trust and trustworthiness cannot occur in equilibrium when all that matters is monetary payoffs. One application is professional (e.g., financial, retail or medical) advice about the most suitable option for a customer. The advisor can either recommend the right option, or mislead the customer with some monetary benefit, such as commissions from product providers, when the customer follows bad advice.[10] At the same time, the customer can choose an outside option that yields the same payoff as the informed decision yet without referring to the advisor. However, the outside option only comes at positive cost γ. This can be interpreted as a costly effort to learn to become informed herself.

Hence, for sufficiently small ϵ the customer's expected payoff from *In* in equilibrium is always smaller than the outside option γ.

[10] An advisor's concern for customers may also derive from liability and threat of prosecution. Still, in many instances unsuitable advice may not be easily detectable and the likelihood of prosecution may remain sufficiently low to act as a sufficient deterrence. It is precisely in these situations where there is the danger that policies could have unwanted consequences by "crowding out" advisors' guilt.

2.2. Psychological payoffs and predictions

We now allow that the advisor is not exclusively motivated by monetary payoffs, but may also be driven by social concerns. Then, if the customer is assumed to be a risk-neutral payoff maximizer, she prefers the strategy (*In, Follow*) over the outside option if

$$\alpha_c \geq \gamma, \tag{1}$$

where α_c is the probability of *Truth* expected by the customer. In what follows, we assume that the customer's beliefs are heterogeneous in the population and *ex ante* unknown to the advisor (see, e.g., Bellemare et al. 2011, for experimental evidence, and Khalmetski et al. 2015 and Attanasi et al. 2015 for models based on a similar assumption).

Our main question is whether the advisor's behavior conditional on *In* may depend on the value γ of the customer's (foregone) outside option. Outcome-based models of social concerns (e.g., Bolton and Ockenfels 2000, Fehr and Schmidt 1999) would not predict any such effect since the mapping of the monetary payoffs of *both* players to the advisor's possible actions remains the same (for different values of γ) once the customer chooses *In*. However, if the advisor's preferences incorporate belief-dependent payoffs, such effects become possible.

2.2.1. Simple guilt

Considering a trust game similar to our Sharing Guilt Game, Battigalli and Dufwenberg (2009) showed that the trustworthiness of the trustee may depend on the (foregone) outside option of the trustor provided that the trustee is guilt averse, i.e., bears psychological costs from disappointing the trustor's *ex ante* payoff expectations. This suggests that something similar can be expected in our context. According to the general model of guilt aversion (Battigalli and Dufwenberg 2007), the psychological cost inflicted on the customer in our game is given by the degree to which her *ex ante* expectations are let down:

$$D_c(s_c, s_a) = \max\{0, E_c^0 - \pi_c(s_c, s_a)\}, \tag{2}$$

where s_c and s_a are the strategies of the customer and the advisor, respectively, E_c^0 is the customer's *ex ante* payoff expectation, and π_c is the realized monetary payoff of the customer. The advisor's guilt is then determined as the share of D_c which could be avoided by the advisor (i.e., which can be attributed to the advisor's choice):

$$G_a(s_c, s_a) = D_c(s_c, s_a) - \min_{\widetilde{s_a}} D_c(s_c, \widetilde{s_a}). \tag{3}$$

Finally, the advisor's utility is

$$U_a(s_c, s_a) = \pi_a(s_c, s_a) - \theta_a G_a(s_c, s_a), \quad (4)$$

where θ_a is the advisor's individual sensitivity to guilt.

Conditional on *In*, the advisor can fully avoid letting down the customer by playing *Truth* so that $\min_{\widetilde{s_a}} D_c(s_c, \widetilde{s_a}) = 0$. If the advisor chooses *Lie* conditional on *In*, his guilt is

$$G_a(In, Lie) = D_c(In, Lie) - 0 = \alpha_c. \quad (5)$$

Denote now by β_a^I the advisor's second-order belief about α_c conditional on observing *In*. The advisor's expected utility from choosing *Lie* after *In* is then

$$E_a[U_a(In, Lie)] = \pi_a^H - \theta_a \beta_a^I. \quad (6)$$

Thus, the advisor prefers to lie if and only if

$$\pi_a^H - \theta_a \beta_a^I \geq \pi_a^L. \quad (7)$$

If $\beta_a^I = 0$, condition (7) always holds so that the advisor prefers lying for any θ_a. If $\beta_a^I > 0$, then by (7) the advisor prefers lying if and only if

$$\theta_a \leq \theta_a^* \equiv \frac{\pi_a^H - \pi_a^L}{\beta_a^I}. \quad (8)$$

Thus, for given $\beta_a^I > 0$, the advisor's optimal strategy can be characterized by the indifferent cutoff type θ_a^* such that only those types, whose individual guilt sensitivity exceeds the cutoff, tell the truth.

Consider finally the effect of the outside option γ on the rate of lying. Since we assume incomplete information about the customer's prior beliefs, the advisor's second-order belief can be updated depending on the customer's choice. Conditional on observing *In*, the advisor's conditional second-order belief about the customer's trust is

$$\beta_a^I := E_a[\alpha_c | In] = E_a[\alpha_c | \alpha_c \geq \gamma], \quad (9)$$

where the last equality follows from optimality of the customer's choice in (1). Thus, β_a^I is increasing in γ (and strictly so if beliefs are sufficiently dispersed): A better foregone outside option of the customer signals her higher expectation of the truth-telling rate. Consequently, given (8), the cut-off level θ_a^* and hence the rate of lying become smaller as γ increases: Higher revealed trust leads to more trustworthiness of a guilt-averse advisor. This is summarized in the following prediction.

Prediction 1 [Simple Guilt]: *Under simple guilt, the advisor becomes less likely to lie as the customer's payoff γ from the foregone outside option increases.*

2.2.2. Shared guilt

Shared guilt introduces a new perspective on the attribution of blame and guilt. If trust gets disappointed, both the advisor's and the customer's actions in our Sharing Guilt Game can be considered as *causes* for the ultimate disappointment of the customer's initial expectations (following the idea that cause is understood as a necessary condition for an outcome to emerge; Lewis 1973). Hence, it is natural to assume that the responsibility for the outcome (in our case, the customer's disappointment) is split between the two causing players.

This hypothesis finds support in psychology research showing that the (feeling of) self-blame depends on how much a person perceives that she could have avoided a negative outcome had she acted differently (Miller and Turnbull 1990, Davis et al. 1996, Mandel 2003). In fact, according to these studies (notably, Davis et al. 1996 and Mandel 2003), increased self-blame, as a result of higher perceived avoidability of the outcome, leads to lower attribution of responsibility to another party (whose actions also contributed to the outcome), and thus to less blaming of this party. Hence, if the latter would anticipate this (as is asserted in our further analysis with respect to the advisor), he or she would suffer less psychological costs related to being blamed, i.e., his or her personal guilt for the outcome might get diffused.

In this section, we introduce a model of shared guilt which captures the idea that the attribution of guilt for disappointing trust is *shared* between players whose choices eventually contributed to this disappointment (including the disappointed player herself). Formally, the psychological cost of the customer from being let down, $D_c(s_c, s_a)$, is a function of *both* the advisor's and the customer's strategy. Hence, in the same way the advisor is treated in (3), we can derive the part of the customer's disappointment which can be attributed to her own behavior, i.e., which could have been avoided had she deviated from her initial plans:

$$G_c(s_c, s_a) = D_c(s_c, s_a) - \min_{\widetilde{s_c}} D_c(\widetilde{s_c}, s_a). \quad (10)$$

We refer to this as "*self-blame*" of the customer. If the customer chooses *In* in equilibrium (so that $\alpha_c \geq \gamma$) and the advisor responds with *Lie*, we obtain, substituting for D_c:

$$\begin{aligned} G_c(In, Lie) &= (\alpha_c - 0) - \min_{\widetilde{s_c}}(\alpha_c - \pi_c(\widetilde{s_c}, Lie)) \\ &= (\alpha_c - 0) - (\alpha_c - \gamma) = \gamma. \end{aligned} \quad (11)$$

Thus, the customer becomes more responsible for her low outcome the less costly the outside option to trusting the advisor.

The remaining part of D_c (i.e., net of self-blame) is then attributed to the advisor, which is referred to as the advisor's "*shared guilt*":

$$\hat{G}_a(s_a) = D_c(s_c, s_a) - G_c(s_c, s_a)$$

$$= D_c(s_c, s_a) - (D_c(s_c, s_a) - \min_{\tilde{s_c}} D_c(\tilde{s_c}, s_a)) = \min_{\tilde{s_c}} D_c(\tilde{s_c}, s_a).[11] \quad (12)$$

Thus, the advisor essentially takes responsibility only for the part of the customer's disappointment that the customer could not have avoided herself (by choosing the outside option).[12] With this modification, the utility of the advisor is, as in the case of simple guilt (equation (4)),

$$U_a(s_c, s_a) = \pi_a(s_c, s_a) - \hat{\theta}_a \hat{G}_a(s_a), \quad (13)$$

where $\hat{\theta}_a$ is the sensitivity parameter with shared guilt.

If the customer prefers to play *In* (so that $\alpha_c \geq \gamma$) and the advisor responds with *Lie*, the advisor's shared guilt is

$$\hat{G}_a(Lie) = \min_{\tilde{s_c}} D_c(\tilde{s_c}, Lie) = \alpha_c - \gamma. \quad (14)$$

The expected advisor's utility in this case is (given that $\beta_a^I > \gamma$ by (9))

$$U_a(Lie) = \pi_a^H - \hat{\theta}_a(\beta_a^I - \gamma), \quad (15)$$

so that the advisor prefers lying over truth-telling if and only if

$$\pi_a^H - \hat{\theta}_a(\beta_a^I - \gamma) \geq \pi_a^L, \quad (16)$$

which is equivalent to

$$\hat{\theta}_a \leq \hat{\theta}_a^* \equiv \frac{\pi_a^H - \pi_a^L}{\beta_a^I - \gamma}. \quad (17)$$

Importantly, in contrast to the cutoff obtained with simple guilt (see (8)), $\hat{\theta}_a^*$ depends on γ for given β_a^I: (17) implies that, ceteris paribus, a higher γ increases the rate of lying. Once the

[11] If the game involved additional players or chance moves, a more general model would add another layer of responsibility by also subtracting from $\hat{G}_a(s_a)$ the part of the customer's disappointment which *neither* the customer nor the advisor could avoid.

[12] As a more general case, suppose that the advisor just assigns a lower weight to guilt corresponding to the customer's self-blame rather than zero weight as in the benchmark case. This model specification is considered in more detail in Appendix A. The qualitative theoretical predictions of the model are robust to this generalization.

customer foregoes a better outside option and still refers to the advisor, this eventually alleviates the latter's (shared) guilt from lying. At the same time, the cutoff still decreases with β_a^I as in the case of simple guilt, which in turn should be positively updated with γ by (9). Thus, the effect of γ on the rate of lying $\hat{\theta}_a^*$ can be decomposed into two effects, where the first one is positive (guilt sharing) and the second one is negative (second-order belief induction):

$$\frac{d\hat{\theta}_a^*}{d\gamma} = \frac{\partial\hat{\theta}_a^*}{\partial\gamma} + \frac{\partial\hat{\theta}_a^*}{\partial\beta_a^I}\frac{\partial\beta_a^I}{\partial\gamma} = \frac{\pi_a^H - \pi_a^L}{(\beta_a^I - \gamma)^2}\left(1 - \frac{\partial\beta_a^I}{\partial\gamma}\right). \quad (18)$$

If the strength of the belief induction effect is sufficiently small ($\frac{\partial\beta_a^I}{\partial\gamma} < 1$), then the effect of guilt sharing dominates: A better outside option foregone by the customer leads to a higher conditional rate of lying.[13]

Prediction 2 [Shared Guilt]: *Under shared guilt, as the customer's foregone outside option increases:*

i) Keeping constant the advisor's belief about the trust of the customer, the advisor is more likely to lie (i.e., he lies also for a lower level of guilt sensitivity).

ii) Taking into account also the induced change in the advisor's second-order belief, the rate of lying increases if $\frac{\partial\beta_a^I}{\partial\gamma} < 1$, *and decreases otherwise.*

Both, simple and shared guilt predict a negative relationship between lying and the advisor's second-order belief. Yet, unlike simple guilt, shared guilt allows for a positive relationship between the value of the foregone outside option and the likelihood of lying. In the next sections, we present results from a test of these predictions in laboratory experiments.

3. Experiment 1: How does the outside option matter for trustworthiness?

In this section, we study how the advisors respond to the customers who forego outside options with different values. The next section presents a robustness check.

[13] Theoretically, one can verify, for instance, that the respective condition $\partial\beta_a^I/\partial\gamma < 1$ holds if α_c is uniformly distributed on [0,1].

	Advisor	*Customer in case of* In	*Customer in case of* Out
The guess is correct	**14**	**14**	**14 – *c***
The guess is incorrect	**17**	**4**	**Max{0,4 – *c*}**

Fig. 2. Payoffs (€) in Experiment 1.

3.1. Experimental design

The experimental game again involved two players (the advisor and the customer) and consisted of the following stages:

(1) The computer randomly chose a natural number from 1 to 100 with each number being equally likely. The advisor got privately informed about the actual number.

(2) The customer decided whether to directly purchase the information about the actual number at cost *c* (*Out*) or to refer to the advisor instead (*In*). In case of purchase, the customer paid the cost, observed the actual number, and the game proceeded to stage 4.

(3) If the customer decided to refer to the advisor, the advisor had to send a message to the customer about the actual number (of the form 'The number is *m*', where *m* is a natural number between 1 and 100).

(4) The customer made a guess about the actual number, and the payoffs were realized.

The final payoffs depended on whether the customer's guess was correct, i.e., matched the computer number (see Fig. 2). The choice and payoff structure is equivalent to the Sharing Guilt Game depicted in Fig. 1 with $\epsilon = 1/99$ (the probability to guess correctly after not following a false advisor's message) and $\gamma = 14 - c$ (with other payoffs being accordingly rescaled).[14] The cost of information *c* varied between 2, 4, 6 and 8, and was deducted from the customer's payoff in case of *Out*.[15]

[14] Because the advisor's action space in the experimental game is not binary as in our Sharing Guilt Game, lying strategies of the advisor in the experimental game may theoretically subsume partially informative strategies, such as sending a truthful number with an added noise term. However, we do not find evidence that subjects were able to coordinate on equilibria with such partially informative communication. In particular, no customer was ever able to guess correctly after receiving a false number. That is, interpreting the advisor's strategy *Lie* in the Sharing Guilt Game to be equivalent to lying in the experimental game is a reasonable approximation, and it is the one we are going to pursue here.

[15] In case of an incorrect guess after purchasing the information (which was never the case in the actual experiment), the customer's payoff was set at 0 if the cost of information exceeded 4 (see Fig. 2).

For reasons described in the next subsection, we implemented an additional treatment, called *No Choice Treatment* (NCT), where stage 2 was omitted. That is, the customer did not have a choice between *In* and *Out* and thus always had to refer to the advisor. Accordingly, the treatments with a possibility to buy information are from now on summarized as *Choice Treatment* (CT).

The game was played for 25 rounds with random rematching of subjects. NCT was played in randomly chosen 3 out of 25 rounds while CT in the remaining 22 rounds (to ensure enough observations where the customer does not buy information in each treatment). The cost of information was chosen randomly each round, with higher frequency for smaller levels in order to (partially) counteract the decline of observations conditional on *In* at lower levels of c.[16] The price of information was made public knowledge between the players at the beginning of a round. At the end of a round, each player observed his/her payoff, while the customer was also shown the actual number chosen by the computer. At the end of the experiment, one round was randomly chosen for payment.

To control for subjects' first- and second-order beliefs, customers were asked about the expected rate of truth-telling in the current round (among advisors who must make the corresponding decision), and advisors were asked to estimate the answer to this question of their currently matched customer. Beliefs were elicited after the players made their decisions, but before they could observe their round payoffs (in particular, after the advisor observed the customer's choice between *In* and *Out*). Beliefs were measured in percentage points. They were incentivized in that subjects were paid additional 4 Euros at the end of the experiment if the corresponding round was chosen for payment and if the actual value (i.e., the actual rate of truth-telling for customers and the matched customer's belief for advisors, respectively) did not differ from their guess by more than 5 percentage points. Players did not receive information whether their beliefs earned an additional payment until the end of the experiment.

The experiment was conducted in the Cologne Laboratory for Economic Research with 236 participants split into 8 sessions. Subjects were recruited with ORSEE (Greiner 2015), and the experiment was computerized with z-Tree (Fischbacher 2007). The average earning was

[16] The within-subject design was used to better control for possible heterogeneity of subjects' preferences in the statistical analysis. As Khalmetski et al. (2015) showed, controlling for such heterogeneity might have crucial implications for the detection of the effect of belief-dependent preferences on prosocial behavior.

near 16 Euro (including a show-up fee of 2.5. Euro), and the experiment lasted for around 1.5 hour.[17] A copy of translated instructions can be found in Appendix B.

3.2. Hypotheses

Simple guilt predicts that as the foregone outside option of the customer gets better, the advisor becomes more trustworthy. This is our Hypothesis 1. Shared guilt is consistent with this hypothesis, but additionally predicts that the customer's choice to forego a more attractive outside option leads to a shift in responsibility for her *ex post* disappointment towards herself. That is, shared guilt can also result in more lying, which constitutes our competing Hypothesis 2.

We note that intention-based reciprocity models (such as Dufwenberg and Kirchsteiger 2004) can be consistent with our Hypothesis 2. The idea is that a higher customer's expectation of truth-telling signaled by foregoing a better outside option leads to a lower estimation of her kind intentions (associated with *In*) by the advisor, and hence to a higher rate of lying. This is why we added NCT to our experiment. The reciprocity models would predict that the rate of lying should be the highest in NCT, because – unlike in CT – there is no scope for positive reciprocity in NCT. Shared guilt predicts the opposite effect: there is no scope for guilt sharing in NCT as the customer takes no choice, so that the advisor's rate of lying should be the lowest in this treatment (conditional on given second-order beliefs). This way, NCT can separate potential explanations based on reciprocity from our shared guilt hypothesis.

Other models that we are aware of, such as outcome-based models, are invariant to changes in c. The same applies to models that assume that truth-telling is driven by a fixed cost of lying, such as in Kartik (2009). Similarly, c plays no role in CT according to action-based reciprocity models (such as Cox et al. 2008, as opposed to intention-based models) once the customer's choice is locked in. No effect of c on advisor behavior is our null hypothesis.

3.3. Results

Fig. 3 shows that the lying rate in CT decreases as the foregone outside option of the customer gets worse, eventually converging to the lying rate in NCT.[18] This is inconsistent with Hypothesis 1, but in line with Hypothesis 2. It suggests that the effect of guilt sharing

[17] After the experiment, the participants had to complete a questionnaire eliciting their age, gender, psychological measures of trust and self-assessment of understanding of the experimental instructions. Control questions aiding basic understanding of the instructions were also asked prior to the experiment.

[18] When reporting average lying rates, the data are first aggregated at the session (matching group) level.

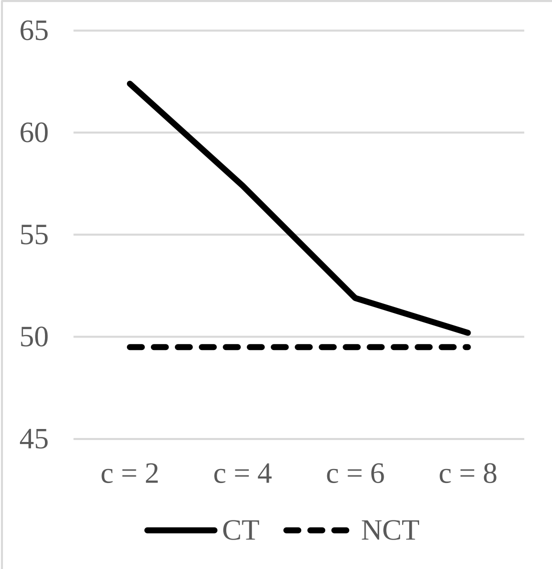

Fig. 3. Lying rate (%) in Experiment 1.

resulting from the customer foregoing a better outside option overcompensates the positive effect on prosocial behavior arising from signaling trust.

Regression analyses strongly confirm the overall picture. Table 1 presents the results of the (random-effects) probit model estimating the determinants of the lying rate. Column 1 shows the basic specification with the exogenous experimental parameters as independent variables. The coefficient on the cost of information c is negative and highly significant, thus supporting the evidence for guilt sharing.

Moreover, the highly significant positive coefficient on the dummy variable for CT supports our hypothesis that the effect is explained by shared guilt rather than reciprocity. Indeed, reciprocity predicts that the advisor considers the customer's choice of *In* as kind, so that the rate of lying in NCT (where no such choice is possible) would be higher than at any value of c in CT. This would then yield a negative coefficient on the dummy variable for CT. Our data rejects this alternative hypothesis.[19]

Next, observe that the advisor's optimal choice in the model of simple guilt (as described in (8)), predicts no effect of the outside option γ once second-order beliefs are controlled for. That is, the total effect of the outside option should then be completely captured by the coefficient on second-order beliefs. In contrast, the model of shared guilt predicts a direct positive effect of γ on the rate of lying for a *given* level of the second-order belief (see (17)). The regression analysis supports shared guilt: The specification in column (2) adds advisors'

[19] The coefficient on the dummy variable for CT estimates the effect on the lying rate resulting from a switch from NCT to a hypothetical counterfactual case of CT with $c = 0$, as c was set to 0 for the observations in the (baseline) treatment NCT.

Table 1. Determinants of the lying rate in Experiment 1, random-effects probit.

Variable	(1)	(2)
CT	1.291***	1.645***
	(0.303)	(0.416)
Cost of information (c)	− 0.178***	− 0.229***
	(0.033)	(0.048)
Round	0.055***	0.047***
	(0.009)	(0.012)
Second-order belief		− 0.033***
		(0.005)
Constant	− 0.745*	0.703*
	(0.433)	(0.372)
Observations	1,032	1,032

Standard errors (clustered at the matching group level) in parentheses. * denotes significance at the 10% level, *** denotes significance at the 1% level.

conditional second-order beliefs to the analysis, which were elicited after the advisor observed the customer's choice between *In* and *Out* in the current round (β_s^I in our model). The effect of these beliefs on the rate of lying is shown to be significantly negative in line with both simple and shared guilt. Yet, at the same time, the effect of the cost of information remains highly negatively significant, as predicted by our model of shared guilt (while being inconsistent with simple guilt).[20]

We conclude that the total effect of the outside option $\gamma = 14 - c$ on the lying rate comprises two effects, as was predicted by (18): the direct effect of guilt sharing and the indirect effect of belief induction, reflected by the corresponding coefficients on c and second-order beliefs, respectively, in column 2 in Table 1. As predicted, due to the negative correlation between c and second-order beliefs, these effects countervail each other. It turns out that, in our data, the effect of guilt sharing is stronger than the opposing effect since the total effect of the cost of information c on the rate of lying is negative (column 1 in Table 1).

We mention that customer behavior is also in line with our model. Table 2 presents the customer's rate of choosing *In* and the corresponding first- and second-order beliefs. Consistent with (1), the rate of *In* is increasing as the cost of choosing the outside option c increases. Also

[20] The regression also shows that there are more lies as advisors become more experienced (the *Round* effect is positive). This complies with the results of Gneezy et al. (2013), who suggest that the phenomenon can be linked to the depletion of self-control.

Table 2. Rate of choosing *In* and conditional beliefs in Experiment 1.

Treatment		Customer's rate of choosing *In, %*	Customer's first-order beliefs conditional on *In, %*	Advisor's second-order beliefs conditional on *In, %*
CT	$c = 2$	7.9	62.7	47.6
	$c = 4$	20.2	61.0	47.7
	$c = 6$	49.4	52.2	43.3
	$c = 8$	65.8	44.1	40.5
NCT		-	40.6	38.9

in line with (1), customers who chose *In* at lower values of *c* are characterized by higher beliefs about the rate of advisors' truth-telling. This is reflected by advisors' second-order beliefs, which tend to follow a similar pattern, though the trend is less pronounced. After getting the message from the advisor, customers set their guess equal to the message in more than 95% of the cases at each level of *c* in CT and in more than 90% of the cases in NCT.

4. Experiment 2: Can self-selection explain our finding?

The main result of Experiment 1 is that in spite of the fact that the customer gets more trusting conditional on foregoing a better outside option, advisors become more likely to lie. Shared guilt suggests that this occurs due to a shift in the attribution of responsibility. One might argue, however, that the effect could be confounded by signaling/self-selection effects induced by the customer's choice of *In* conditional on a given outside option. For instance, asking the advisor for advice even when there would have been a cheap alternative to get informed might signal specific individual characteristics of a given customer, e.g., her low risk or loss aversion, which in turn could affect advisors' behavior towards this customer, and probably even so in the same direction as shared guilt. Because this cannot be excluded *ex ante*, we conducted Experiment 2, which allows to disentangle *any* potential selection/signaling effects of the customer's choice from the effect of a responsibility shift predicted by shared guilt.

4.1. Experimental design

The experimental game is the same as before, and is played repeatedly for 26 rounds with random rematching of players. Also as before, we had rounds with no customer's option to buy information, and rounds that included such an option, with the same possible values of *c*. Yet, this time the treatments were played in a new, specific order. A round with no option to buy information always alternated with a round with such an option. At the beginning of each round with *no option* to buy information the advisor was informed about the choice of the currently matched customer in the *previous* round (while being reminded about the cost of information in that round). This is why we denote the rounds with no option to buy information as Previous Choice Treatment (PCT). The rounds *with an option* to buy information are denoted as Current Choice Treatment (CCT). CCT in Experiment 2 is equivalent to CT in Experiment 1 in terms of both (material) game and information structure, thus serving as a control treatment.

In each session, there were 13 rounds in CCT and 13 rounds in PCT. No subject played with the same partner in two consecutive rounds, which was made public knowledge. This ensured that the customer's choice in the previous round may affect the current advisor's behavior only via signaling of information about preferences/beliefs of the customer. The experiment was conducted in the Cologne Laboratory for Economic Research with 136 participants, split into 5 sessions. In each session subjects were divided into 2 independent matching groups. A copy of translated instructions can be found in Appendix B.

4.2. Hypotheses

As in CT, guilt sharing predicts an increase in the lying rate with the customer's (foregone) outside option in CCT. On the other side, PCT leaves no scope for self-blame of the customer (in the sense of (10)), since there is no current option to buy information, and hence the customer cannot (retrospectively) reduce her disappointment in case of lying by taking another choice.[21]

At the same time, all potential signaling effects of the customer's observed choice in PCT remain the same as in CCT. Precisely, if the advisor reacts to the customer's (fixed) preferences as revealed by her behavior, it should not matter whether this information stems from her behavior in the current or that in the previous round. Therefore, if the previously established positive correlation between the customer's unchosen outside option and the conditional lying rate was driven by some kind of signaling/selection effects, it should be observed in both PCT and CCT. At the same time, if this correlation was rather caused by the sharing of guilt induced by the (current) customer's choice of *In*, then this effect should not be observed in PCT while it should arise in CCT.

[21] The customer can only slightly increase her expected payoff from 4 to 4.1 by not following the false message, as the probability to guess correctly at random is 1/99. This still leaves almost all of her disappointment attributed to the advisor according to (12).

Table 3. Determinants of the lying rate in Experiment 2, random-effects probit.

Variable	*CCT*		*PCT*	
	(1)	(2)	(3)	(4)
Cost of information (c)	– 0.199**	– 0.209**	0.038	0.046
	(0.100)	(0.091)	(0.057)	(0.069)
Round	0.057***	0.053**	0.037	0.035
	(0.015)	(0.024)	(0.030)	(0.025)
Second-order belief		– 0.038***		– 0.039***
		(0.008)		(0.007)
Constant	– 0.509	1.802**	– 1.301	0.814
	(0.578)	(0.850)	(0.887)	(0.889)
Observations	277	277	277	277

Standard errors (clustered at the matching group level) in parentheses. ** denotes significance at the 5% level, *** denotes significance at the 1% level.

4.3. Results

The results for the lying rate in CCT replicate the previously established effect of guilt sharing, as manifested by the regression results in Table 3 (columns 1 and 2). In particular, the coefficient on the cost of information *c* is significantly negative both with and without control for second-order beliefs. In stark contrast, the customer's outside option, foregone in the previous round, has no significant effect on the lying rate in PCT, with the effect, if at all, pointing even in the opposite direction (see Table 3, columns 3 and 4).

The results for the customer's behavior in CCT largely follow the observations in CT in Experiment 1 (see Table 4). The rate of *In* increases with the cost of information, while the conditional first-order beliefs of (self-)selected customers decrease. Advisors' second-order beliefs also tend to follow a decreasing pattern with respect to *c* in PCT, and to a lesser extent in CCT (see Table 4).

We conclude that the previously established positive effect of the customer's unchosen outside option on the rate of lying was not driven by signaling/selection effects. Instead, this effect reveals itself only if the customer makes an explicit (trusting) choice in the current interaction, which, as our theory suggests, allows to reallocate responsibility for the final outcome to the customer.

Table 4. Rate of choosing *In* and conditional beliefs in Experiment 2.

Cost of information	Customer's rate of choosing *In*, %	Customer's first-order beliefs conditional on *In*, %	Advisor's second-order beliefs conditional on *In*, %	
			CCT	*PCT*
$c = 2$	12.5	73.2	54.2	71.0
$c = 4$	31.3	66.6	55.4	52.8
$c = 6$	49.3	59.7	45.7	53.5
$c = 8$	69.6	60.1	50.6	48.6

5. Conclusion

Our study provides empirical support for our model of guilt sharing in strategic communication. Our laboratory game allows advisors to assign a part of their responsibility for the final outcome to customers who, in turn, could have avoided being deceived by acquiring information on their own. Making the customers' information acquisition less costly led to more responsibility shifting and less prosocial behavior of the advisors. The effect cannot be explained by standard models of reciprocity, belief- or outcome-dependent preferences, but is consistent with a new model of shared causal attribution of guilt, which builds on earlier work on guilt aversion.

Our study suggests that, as the information age makes it easier for people to get informed, say, about product quality, competitors' prices or seller trustworthiness, people are made more accountable for their choices, which may reduce others' responsibilities to take care. At the same time, our study may inform firm policies and consumer protection regulation, such as in Internet markets or the market for financial advice, by pointing out unintended negative consequences of reducing consumers' need to rely on human experts, which might mitigate the latter's responsibility to behave in a trustworthy manner.

Further research can be done to accommodate more complex structural characteristics of advice situations. For instance, one can investigate whether the effect of guilt sharing is retained by partial information acquisition on the side of the customer, after which he still has to consult an advisor. Another relevant question is whether the effect is robust to reputational concerns of the advisor in repeated interactions. From the methodological perspective, it remains an open question whether our concept of shared guilt is applicable to a broader class of trust games, e.g., to the games analyzed by Charness and Rabin (2002). Besides, it would be worthwhile to explore whether trustors anticipate the potential responsibility shift induced by foregoing a valuable outside option, and adjust their behavior accordingly (for instance, by being reluctant to disclose their prior outside option to the trustee).

Our concept of shared guilt suggests a new perspective on the effect of increased informational transparency in the economy, and on how responsibility and blame can be shared in strategic interactions. Given its success in organizing our and others' laboratory evidence, our framework might prove to be useful for developing further applied models involving communication and trust.

References

Ahmad, F., Hudak, P.L., Bercovitz, K., Hollenberg, E., Levinson, W. (2006): "Are Physicians Ready for Patients with Internet-Based Health Information?" *Journal of Medical Internet Research*, 8(3), e22.

Akerlof, G.A. (1970): "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, 84(3), 488–500.

Anagol, S., Kim, H.H. (2012): "The Impact of Shrouded Fees: Evidence from a Natural Experiment in the Indian Mutual Funds Market," *American Economic Review*, 102(1), 576–593.

Attanasi, G., Battigalli, P., Manzoni, E. (2015): "Incomplete-Information Models of Guilt Aversion in the Trust Game," *Management Science*, 62(3), 648–667.

Bakos, J.Y. (1997): "Reducing Buyer Search Costs: Implications for Electronic Marketplaces," *Management Science*, 43(12), 1676–1692.

Bartling, B., Fischbacher, U. (2012): "Shifting the Blame: On Delegation and Responsibility," *Review of Economic Studies*, 79, 67–87.

Battigalli, P., Dufwenberg, M. (2007): "Guilt in Games," *American Economic Review*, 97, 170–176.

Battigalli, P., Dufwenberg, M. (2009): "Dynamic Psychological Games," *Journal of Economic Theory*, 141, 1–35.

Beck, A., Kerschbamer, R., Qiu, J., Sutter, M. (2013): "Shaping Beliefs in Experimental Markets for Expert Services: Guilt Aversion and the Impact of Promises and Money-Burning Options," *Games and Economic Behavior*, 81, 145–164.

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 Econometrics*, 26, 437–453.

Bolton, G.E., Ockenfels, A. (2000): "ERC: A Theory of Equity, Reciprocity, and Competition," *American Economic Review*, 90, 166–193.

Bolton, G., Greiner, B., Ockenfels, A. (2013): "Engineering Trust: Reciprocity in the Production of Reputation Information," *Management Science*, 59(2), 265–285.

Brandts, J., Fatas, E., Haruvy, E., Lagos, F. (2015): "The Impact of Relative Position and Returns on Sacrifice and Reciprocity: An Experimental Study Using Individual Decisions," *Social Choice and Welfare*, 45(3), 489–511.

Cain, D.M., Loewenstein, G., Moore, D.A. (2005): "The Dirt on Coming Clean: Perverse Effects of Disclosing Conflicts of Interest," *Journal of Legal Studies*, 34(1), 1–25.

Cain, D.M., Loewenstein, G., Moore, D.A. (2011): "When Sunlight Fails to Disinfect: Understanding the Perverse Effects of Disclosing Conflicts of Interest," *Journal of Consumer Research*, 37(5), 836–857.

Charness, G. (2000): "Responsibility and Effort in an Experimental Labor Market," *Journal of Economic Behavior and Organization*, 42(3), 375–384.

Charness, G., Cobo-Reyes, R., Jiménez, N., Lacomba, J.A., Lagos, F. (2012): "The Hidden Advantage of Delegation: Pareto Improvements in a Gift Exchange Game," *American Economic Review*, 102(5), 2358–2379.

Charness, G., Dufwenberg, M. (2006): "Promises and Partnership," *Econometrica*, 74, 1579–1601.

Charness, G., Rabin, M. (2002): "Understanding Social Preferences with Simple Tests," *Quarterly Journal of Economics*, 117(3), 817–869.

Coffman, L.C. (2011): "Intermediation Reduces Punishment (and Reward)," *American Economic Journal: Microeconomics*, 3(4), 77–106.

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.

Davis, C.G., Lehman, D.R., Silver, R.C., Wortman, C.B., Ellard, J.H. (1996): "Self-Blame Following a Traumatic Event: The Role of Perceived Avoidability," *Personality and Social Psychology Bulletin*, 22, 557–567.

Duarte, F, Hastings, J.S. (2012): "Fettered Consumers and Sophisticated Firms: Evidence from Mexico's Privatized Social Security Market," working paper no. w18582, National Bureau of Economic Research.

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(2), 268–298.

Erat, S., Gneezy, U. (2012): "White Lies," *Management Science*, 58(4), 723–733.

Fehr, E., Schmidt, K.M. (1999): "A Theory of Fairness, Competition and Cooperation," *Quarterly Journal of Economics*, 114, 817–868.

Fershtman, C., Gneezy, U. (2001): "Strategic Delegation: An Experiment," *RAND Journal of Economics*, 32(2), 352–368.

Fischbacher, U. (2007): "z-Tree: Zurich Toolbox For Ready-Made Economic Experiments," *Experimental Economics*, 10, 171−178.

Garofalo, O., Rott, C. "Shifting Blame? Experimental Evidence on Delegating Communication," *Management Science*, forthcoming.

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

Gneezy, U. (2005): "Deception: The Role of Consequences," *American Economic Review*, 95, 384–394.

Gneezy, U., Rockenbach, B., Serra-Garcia, M. (2013): "Measuring Lying Aversion," *Journal of Economic Behavior and Organization*, 93, 293–300.

Golan, E., Kuchler, F., Mitchell, L., Greene, C., Jessup, A. (2001): “Economics of Food Labeling,” *Journal of Consumer Policy*, 24(2), 117–184.

Goldmanis, M., Hortaçsu, A., Syverson, C., Emre, Ö. (2010): “E-Commerce and the Market Structure of Retail Industries,” *Economic Journal*, 120(545), 651–682.

Greiner, B. (2015): “Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE,” *Journal of the Economic Science Association*, 1(1), 114–125.

Hamman, J.R., Loewenstein, G., Weber, R.A. (2010): “Self-Interest through Delegation: An Additional Rationale for The Principal-Agent Relationship,” *American Economic Review*, 100(4), 1826–1846.

Inderst, R., Ottaviani, M. (2012): “Competition through Commissions and Kickbacks,” *American Economic Review*, 102(2), 780–809.

Kartik, N. (2009): “Strategic Communication with Lying Costs,” *Review of Economic Studies*, 76, 1359–1395.

Kaufmann, D., Bellver, A. (2005): “Transparenting Transparency: Initial Empirics and Policy Applications,” working paper, World Bank.

Khalmetski, K. (2016): “Testing Guilt Aversion with an Exogenous Shift in Beliefs,” *Games and Economic Behavior*, 97, 110–119.

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

Kozup, J., Hogarth, J.M. (2008): “Financial Literacy, Public Policy, and Consumers’ Self-Protection – More Questions, Fewer Answers,” *Journal of Consumer Affairs*, 42(2), 127–136.

Lacko, J., Pappalardo, J. (2007): “Improving Consumer Mortgage Disclosures: An Empirical Assessment of Current and Prototype Disclosure Forms,” Federal Trade Commission Bureau of Economics Staff Report, Washington, DC.

Lewis, D. (1973). “Causation”, *Journal of Philosophy*, 70(17), 556–67.

Mandel, D. (2003): “Counterfactuals, Emotions, and Context,” *Cognition and Emotion*, 17(1), 139–159.

Miller, D.T., Turnbull, W. (1990): “The Counterfactual Fallacy: Confusing What Might Have Been with What Ought to Have Been,” *Social Justice Research*, 4(1), 1–19.

Oexl, R., Grossman, Z.J. (2013): “Shifting the Blame to a Powerless Intermediary,” *Experimental Economics*, 16(3), 306–312.

Reuben, E., Sapienza, P., Zingales, L. (2009): “Is Mistrust Self–Fulfilling?” *Economics Letters*, 104, 89–91.

Servátka, M., Vadovič, R. (2009): “Unequal Outside Options in the Lost Wallet Game,” *Economics Bulletin*, 29(4), 2870–2883.

Stiglitz, J.E., Weiss, A. (1981): “Credit Rationing in Markets with Imperfect Information,” *American Economic Review*, 71(3), 393–410.

Woods, D., Servátka, M. (2016): Testing Psychological Forward Induction and the Updating of Beliefs in the Lost Wallet Game,” *Journal of Economic Psychology*, 56, 116–125.

Appendix A: Generalized preferences

One can show that the theoretical predictions of the model of shared guilt are robust to a generalization of preferences. Specifically, in our main specification, we assumed that the advisor does not feel guilt for the share of the customer’s disappointment which could be avoided by the customer. One can think of a more general specification where the advisor assigns a lower weight to this share of the customer’s disappointment instead of ignoring it completely. In this case, the (modified) shared guilt of the advisor can be written as

$$\begin{aligned}\hat{G}_a(s_a) &= (D_c(s_c, s_a) - G_c(s_c, s_a)) + \omega G_c(s_c, s_a) \\ &= \omega D_c(s_c, s_a) + (1-\omega) \min_{\widetilde{s_c}} D_c(\widetilde{s_c}, s_a), \qquad (19)\end{aligned}$$

where $\omega < 1$ is the degree of responsibility felt by the advisor for the part of the customer’s disappointment corresponding to her self-blame. In particular, $\omega = 0$ corresponds to shared guilt while $\omega = 1$ to simple guilt.

Then, given (19), the advisor’s expected utility from lying after *In* is

$$U_a(In, Lie) = \pi_a^H - \hat{\theta}_a(\omega\beta_a^I + (1-\omega)(\beta_a^I - \gamma)), \qquad (20)$$

which gives rise to the cutoff

$$\hat{\theta}_a^* = \frac{\pi_a^H - \pi_a^L}{\beta_a^I - (1-\omega)\gamma}. \quad (21)$$

In this case, $\hat{\theta}_a^*$ increases with γ if

$$\frac{\partial \beta_a^I}{\partial \gamma} < 1 - \omega. \quad (22)$$

This generalizes the condition on the intensity of second-order belief updating required to yield the positive correlation between the value of the customer's foregone outside option and the lying rate.

Appendix B: Experimental instructions (translated from German)

Experiment 1

Welcome to our experiment!

You are now participating in an experiment in which you can earn money. Your total payoff depends on your own decisions and those of the other participants. Please refrain from now on from talking and looking at other participants' screens. If you have any questions, please raise your hand. We will come to your place and answer your question as soon as possible. During the experiment you will interact with other participants. The identity of the other participants will not be revealed to you. Likewise, your identity will not be revealed to the other participants. These instructions are identical for all participants.

Before the experiment begins, you will be assigned the role of either the "advisor" or the "customer". This role will be retained during the entire experiment.

The experiment consists of 25 rounds. In every round, each advisor is matched to a new customer. The matching in every round is random.

In every round, the computer randomly draws a "secret" number between 1 and 100 for each participant pair. Each number has the same chance of being drawn. The advisor learns the drawn number at the beginning of each round but the customer does not.

The customer's task is to guess the "secret" number. The payoffs of both players depend on whether the customer's guess is correct or false, and are as follows (in "game points"):

	Points Advisor	Points Customer
The customer's guess is *correct*	**14**	**14**
The customer's guess is *false*	**17**	**4**

In some rounds, the customer can acquire information about the secret number only from his advisor. In the other rounds, the customer has an additional option to acquire information about the secret number by himself (without his advisor). However, this information is not free and costs to the customer 2, 4, 6 or 8 points. Both participants are informed at the beginning of each round whether or not the customer can acquire information about the secret number and how much it costs. If the customer decides to acquire information, the actual secret number is revealed to him. The information costs are then subtracted from his payoff for the round. If the payoff is not sufficient to cover the information costs, his final payoff for the round is 0 points.

If the customer decides to proceed without buying information from the computer or if this option is not available, the advisor must send to the customer a message in the following form:

"The secret number is …"

The advisor can transmit any possible number independently of the actual number.

After the customer has bought information or has received a message from his advisor, he has to make a guess. At the very end of the round, each participant learns how many points he and his fellow player received in this round. Additionally, the customer learns the actual number which was observed by the advisor.

At the end of the entire experiment, one of the 25 rounds is randomly chosen for payment. The same round is chosen for all participants. *For the payout, 1 point equals to one euro.* Additionally, you always receive 2.50 euro for showing up.

After signing the receipt, you will receive your payoff in cash.

At the end of the experiment, we will also ask you to answer a short questionnaire on your computer.

Please click the button "Done" once you have read and understood the instructions.

Experiment 2

Welcome to our experiment!

You are now participating in an experiment in which you can earn money. Your total payoff depends on your own decisions and those of the other participants. Please refrain from now on from talking and looking at other participants' screens. If you have any questions, please raise your hand. We will come to your place and answer your question as soon as possible. During the experiment you will interact with other participants. The identity of the other participants will not be revealed to you. Likewise, your identity will not be revealed to the other participants. These instructions are identical for all participants.

Before the experiment begins, you will be assigned the role of either the “advisor” or the “customer”. This role will be retained during the entire experiment.

The experiment consists of 26 rounds. In every round, each advisor is matched to a new customer. The matching in every round is random. The matching ensures that the same advisor and customer never interact with each other in two successive rounds.

In every round, the computer randomly draws a “secret” number between 1 and 100 for each participant pair. Each number has the same chance of being drawn. The advisor learns the drawn number at the beginning of each round but the customer does not.

The customer’s task is to guess the “secret” number. The payoffs of both players depend on whether the customer’s guess is correct or false, and are as follows (in “game points”):

	Points Advisor	Points Customer
The customer’s guess is *correct*	**14**	**14**
The customer’s guess is *false*	**17**	**4**

In some rounds, the customer can acquire information about the secret number only from his advisor. In the other rounds, the customer has an additional option to acquire information about the secret number by himself (without his advisor). However, this information is not free and costs to the customer 2, 4, 6 or 8 points. Both participants are informed at the beginning of each round whether or not the customer can acquire information about the secret number and how much it costs. If the customer decides to acquire information, the actual secret number is revealed to him. The information costs are then subtracted from his payoff for the round. If the payoff is not sufficient to cover the information costs, his final payoff for the round is 0 points.

The advisor is immediately notified whether his customer has bought information. Additionally, the advisor learns whether the customer bought information in the previous round (while he was assigned to a different advisor).

If the customer decides to proceed without buying information from the computer or if this option is not available, the advisor must send to the customer a message in the following form:

“The secret number is …”

The advisor can transmit any possible number independently of the actual number.

After the customer has bought information or has received a message from his advisor, he has to make a guess. At the very end of the round, each participant learns how many points he and his fellow player received in this round. Additionally, the customer learns the actual number which was observed by the advisor.

At the end of the entire experiment, one of the 26 rounds is randomly chosen for payment. The same round is chosen for all participants. *For the payout, 1 point equals to one euro.* Additionally, you always receive 2.50 euro for showing up.

After signing the receipt, you will receive your payoff in cash.

At the end of the experiment, we will also ask you to answer a short questionnaire on your computer.

Please click the button “Done” once you have read and understood the instructions.