Judgment and Decision Making, Vol. 12, No. 4, July 2017, pp. 408–414

The value of vulnerability: The transformative capacity of risky trust

Luigino Bruni* Fabio Tufano†

Abstract

In an experimental gift-exchange game, we explore the transformative capacity of vulnerable trust, which we define as trusting untrustworthy players when their untrustworthiness is common knowledge between co-players. In our experiment, there are two treatments: the "Information" treatment and the "No-Information" treatment in which we respectively disclose or not information about trustees' trustworthiness. Our laboratory evidence consistently supports the transformative capacity of trustors' vulnerable trust, which generates higher transfers, more trustworthiness and increased reciprocity by untrustworthy trustees.

Keywords: experiment, gift-exchange game, organization, trust, vulnerability

1 Introduction

Behavioural and social scientists have been increasingly studying trust and its properties (e.g., Balliet et al., 2013; Fehr, 2009; Johnson & Mislin, 2011). Still, much needs to be understood about trust, particularly in non-enforceable, personalised interactions. The study reported here investigates experimentally trustees' response when the *intentional* vulnerability of the trustor is both manifestly salient and clearly dependent upon the trustee's revealed trustworthiness.

That trust – when not purely self-interested and instrumental – involves vulnerability is acknowledged in the interdisciplinary literature on trust (e.g., Rousseau et al., 1998; Schoorman et al., 2007). Vulnerability, however, is often interpreted only as unintentional "exposure" to other people's action or events, normally due to lack of resources, rights, capabilities, empowerment or freedom.[1] The development of human wellbeing and dignity is usually measured in terms of reduction or elimination of this unintentional vulnerability. At the same time, some philosophers and social scientists claim also for a purposeful vulnerability, related to the inherent fragility associated to good life (e.g., Nussbaum, 1986).

Vulnerability pervades the most diverse social interactions (ranging from education to management; from family to public life) whenever one party, the trustor, intentionally entrust her "fate" to another party, the trustee, who is free to behave in ways that may be beneficial but even in ways that may be harmful for the trustor. This could be the case, for instance, of a grassroots manager trusting a selfish young footballer to play for the team in the cup final. The young footballer may betray her trust and use the cup final to showcase his skills, with little consideration of the team's objectives; or, given his history of selfish player and aware of the intentional vulnerability of the manager's trust, the footballer may change behaviour and reciprocate her trust by playing to benefit the team.

Everyday life shows abundant evidence that without this purposeful, intentional, vulnerability, human life does not flourish fully, and organizations do not fulfil entirely their potential.

By means of a gift-exchange experiment, in this paper we study trustee's response to trustor's intentional vulnerability and, consequently, the transformative capacity of vulnerable trust, which we conceptualise as the possibility that trusting untrustworthy individuals may change their responses from untrustworthy responses to trustworthy ones. To the best of our knowledge, no experiment has been conducted so far to investigate the effects of trusting untrustworthy individuals, when this specific vulnerability is common knowledge and is made salient by its manifest intentionality. In what follows, we refer to the trustor's risk to be potentially betrayed by the trustee, who proved to be untrustworthy (i.e., by sending back less than what received) in a recent interaction with a third person, as trustor's vulnerability. Our general concern is: Does vulnerable trust increase trustees' transfers? Does it transform trustees' attitudes by making their behaviour more trustworthy and reciprocal? Our laboratory evidence says unambiguously *yes* to both questions. Intentional vulnerability of trust shows sizeably increases in trustee's transfers.

Part of the research was done when Bruni and Tufano were at the University of Milano-Bicocca (Italy). The authors thank Irene Brundia for help and suggestions at the early stage of the design, and for lab assistance. We thank also Luca Stanca for his inputs and support especially in the initial phases of the study. Research funding from the Center for Interdisciplinary Studies in Economics, Psychology and Social Sciences (University of Milano-Bicocca) is gratefully acknowledged.

*Dipartimento di Scienze economiche, politiche e delle lingue moderne, LUMSA, Via Pompeo Magno 22, 00192 Roma, Italy. E-mail: l.bruni@lumsa.i.

†CeDEx, School of Economics, University of Nottingham, University Park, Nottingham NG7 2RD, UK. E-mail: fabio.tufano@nottingham.ac.uk.

[1] On the concept of "exposure" see Pelligra (2007) and also Bruni (2012).

https://doi.org/10.1017/S1930297500006276 Published online by Cambridge University Press

Trustees' behaviour becomes also more trustworthy and reciprocal when vulnerability is made salient. We discuss the policy implications of our results in significant domains of social and economic life such as management and education, where personalised interactions are characterised by non-enforceable trust.

2 Trust, Risk and Vulnerability

Berg et al. (1995), Pillutla et al. (2003), Malhotra (2004) and Strassmair (2009) dealt with issues similar to our own by explicitly investigating in an experimental setting the role of trustors' risk exposure with regard to trustees' behaviour.[2] These studies did not find any significant effect of the trustor's exposure on the trustee's behaviour. These results depend on critical features of the experimental designs, which – we believe – were not purposely built to investigate what we refer to as the transformative effect of vulnerability on trustee's behaviour.

In particular, in Berg et al. (1995) and Pillutla et al. (2003), whenever trustors took higher risks by sending larger portions of their endowment, they provided greater benefit to trustees as a consequence. Thus, there is no possibility to know whether trustees reciprocated because they valuated the risks trustors had undertaken, or for distributional reasons – or for both.

Malhotra's (2004) study is the closest to our own study, with a specific acknowledgement of the role of the trustor's risk. He found no significant impact of the trustor's risk on the trustee's trustworthiness, which is instead significantly affected by the benefit provided to them by the trustor. In contrast with previous studies, Malhotra's (2004) experimental design maintained a clear separation between the effect of the trustor's risk and the trustee's benefit, but still the trustor's intentional vulnerability was not manifestly salient and clearly dependent upon the trustee's revealed trustworthiness. In fact, in his study the only dimension of risk exposure experimentally manipulated was the variation in the material payoffs of the trustor's outside option. We claim that this strategy is not able to analyse the role of trustor's intentional vulnerability, i.e. the specific risk of being betrayed inherent to trustor's interaction with a given (untrustworthy) trustee.

Strassmair (2009) studied the effect of trustors' expected future rewards on trustees' reciprocal responses. To this aim, she experimentally varied across treatments the probability for the trustee to make a return transfer. In all her experimental treatments, the probability of a return transfer was made common knowledge when subjects were instructed. In Strassmair's (2009) "low" treatment, the probability for the trustees of making their return transfer was 50%; in her "high" treatment, the respective probability was 80%. Therefore, in the low treatment the trustee were expected to perceive the trustor as kinder than in the high treatment, ceteris paribus, and therefore they were expected to return more in the former than in the later treatment whenever asked to make a decision. The results, however, did not show any significant difference across treatments: in fact, trustees did not condition their transfers on trustors' expected future rewards. Therefore, she suggested that trustees are insensitive to the specific risk faced by the trustors and, consequently, how risky trustors' trust was. But, again, also in Starssmair's experiment the specific dimension of intentional vulnerability was not salient enough, due to the mediational role played by the probability of making a return transfer, and did not depend upon the trustee's revealed trustworthiness.

In fact, the contribution of our paper is to explore the role of *intentional vulnerability*. The vulnerability of trust is explained as a disposition of the trustor to accept the risk to be *intentionally* betrayed by the trustee (Baier, 1986). This disposition emerges only in context of human relationships in which the presence of people and their intentions – rather than other elements of the decision context – explains the possibility to feel betrayed rather than the mere possibility to be disappointed.

Thus, we consider the presence of people and their intentions as a first necessary condition for the vulnerability of trust as we intend it. However, it does not constitute a sufficient condition to explain the emergence and the role of intentional vulnerability. As also Holton (2004) underlines, a person could choose to undertake actions based on trust without taking the risk to be betrayed (e.g., when interacting with an absolutely trustworthy person), that is, without any particular vulnerability. This leads to what we consider a second necessary condition for the vulnerability of trust: that is, the risk of trusting depends on the trustee's revealed level of trustworthiness.

Upon maintaining the first condition in each and every experimental treatment as in previous related studies (e.g., Blount, 2005; Falk et al. 2008; Stanca et al. 2009; Stanca, 2010), we design our experiment to explore the effect of the second condition once the trustee's revealed level of trustworthiness has been made manifestly salient. We hypothesise that, when the second condition also holds, trustors' *vulnerability* may have a transformative effect on the response of untrustworthy trustees.[3] This is what we investigate by means of our experiment as detailed below.

[2]Cialdini (1993) and Reagan (1971) dealt with risk and reciprocity, but the issue of vulnerability was not part of their analysis.

[3]For instance, we speculate that when the second condition holds trustors' vulnerability may prompt trustees to regard norms of altruism, reciprocity, and fairness as relevant for the circumstances they are in, and behave accordingly rather than following the norm of self-interest (e.g., Miller, 1999; Ratner & Miller, 2001). However, it could also be the case that behaviour in line with norms of altruism, reciprocity and fairness is the intuitive behaviour triggered without deliberation (e.g., Rand, 2016, on cooperation as resulting from more intuitive or deliberative processes).

https://doi.org/10.1017/S1930297500006276 Published online by Cambridge University Press

3 Experimental Design and Procedures

We employ a two-player symmetric gift-exchange game (Stanca et al., 2009). Both players receive an initial endowment of 5 tokens each. In line with the literature on trust (e.g., Bohnet, 2008), we refer to the first mover as the trustor and the second mover as the trustee. In Stage 1, the trustor decides how many of her 5 tokens (only integers could be disposed) to send to the trustee. Then, the x tokens sent by the trustor are multiplied by 3 by the experimenter. Therefore, the trustee receives $3x$. In Stage 2, the trustee decides how many of his 5 tokens (only integers could be disposed) to send to the trustor. Then, the y tokens sent by the trustee are multiplied by 3 by the experimenter. Thus, the trustor receives $3y$. In summary, the trustor's final payoff of the game is $5 - x + 3y$, while the trustee's final payoff is $5 + 3x - y$.

Within each experimental session, the game was played three times. We refer to those three times as Game 1, 2 and 3, respectively. Players learned about the games step by step. Players' roles – namely either trustor or trustee – were fixed across games. A stranger matching protocol was in place: that is, Game 1–3 were each played with a different co-player. Only in Stage 2 of Game 3 we applied a variant of the strategy method similar to the one implemented in Fischbacher et al. (2001): that is, the trustee in such a stage had to make a *set of six conditional decisions* (i.e., one per possible number of tokens they could receive from the trustor they were paired with) without knowing how many tokens the trustor actually sent. In all other stages, the decision method was always applied entailing a *single unconditional decision* about how many tokens from the initial endowment a player sent to their co-player.

One of the three games, 1–3, was randomly selected for payment. For trustors, the single unconditional decision in the selected game was used to determine the game payoffs relevant for payment. For trustees, if either Game 1 or 2 was selected the single unconditional decision was used to determine the game payoffs relevant for payment, whereas if Game 3 was selected the conditional decision corresponding to their trustor's unconditional decision was used. Experimental earnings were obtained by converting the payoffs of the selected game in euros (exchange rate: 2 tokens = 1 euro), plus 5 euros as show-up fee.

There are two treatments in the experiment: the Information treatment (I-treatment) and No-Information treatment (N-treatment). The experimental manipulation between treatments is the disclosure or not of information about the trustee's choice in Game 1. In fact, in Game 2 and 3 of the I-treatment, trustors are informed whether their co-player made either a "trustworthy" or an "trustworthy" choice in Game 1, while trustees were made aware that their trustor co-players were informed whether they made either a "trustworthy" or a "trustworthy" choice in Game 1. By contrast, in Game 2 and 3 of the N-treatment, no information about Game 1 was disclosed. To effectively and swiftly inform trustors about trustees' revealed trustworthiness, a trustee's choice in Game 1 was labelled as "trustworthy" (vs. "untrustworthy") if they sent to their co-player a number of token larger than or equal to (vs. lower than) the tokens they received. (The experimental instructions rather than "trustworthy" used the more neutral Italian term "equo" – or its negation – that could be more closely translated in English with the word "fair"). The two treatments were identical in all other respects.

The experiment started with instructions read aloud by the experimenters to set ground rules. Then, subjects were led step by step by computerized instructions in z-Tree (Fischbacher, 2007). After going through Games 1–3, subjects learned the game randomly selected for payment, their choice, their opponent's choice and their earnings. The experiment ended with a standard background questionnaire.

Two-hundred eight students (of whom 59.62 percent enrolled in undergraduate degrees) drawn from a range of academic disciplines (with Business and Economics summing up respectively to 50.96 percent and 17.79 percent of the whole sample) participated in our experiment, which took place at the Experimental Economics Laboratory of the University of Milano-Bicocca (Italy) and lasted on average one hour. Subjects were paid individually and anonymously at the end of each experimental session.[4]

4 Predictions

The present study focuses on the behavioural implications of the vulnerability of trust. To explore those implications, it is necessary to concentrate the attention on the trustee. Therefore, in what follows, the predictions are stated with regard to trustees' behaviour.

Upon assuming that players are purely self-interested and this is common knowledge, the trustee who is at the game terminal node will always send zero tokens to the trustor. By backward induction, the trustor will rationally choose to not send any token to the trustee. Therefore, the standard equilibrium prediction is that both players will send zero tokens. However, we do not expect many players to behave this way.

More importantly in the present case, if trustee's preferences show concerns for the vulnerability of trust, the amount of tokens sent back by them should be higher when the trustor's vulnerability is salient.

[4]All subjects received the total sum of the actual earnings from the experiment as described in the main text plus a € 5.00 show-up fee. Total payments ranged between € 5.00 and € 15.00 with an average payment equal to € 9.40 (standard deviation of € 3.12).

https://doi.org/10.1017/S1930297500006276 Published online by Cambridge University Press

Table 1: Summary statistics of token transfers by treatment, game and type of player.

	Game 1 Mean (s.d.)	Game 2 Mean (s.d.)	Game 3 Mean (s.d.)	N
N-treatment				
Trustor's transfer	3.077 (1.702)	3.192 (1.783)	2.135 (1.961)	52
Trustee's transfer	1.885 (1.843)	1.885 (1.916)	1.474 (0.996)	52
I-treatment				
Trustor's transfer	2.846 (1.564)	2.365 (1.794)	2.731 (2.097)	52
Trustee's transfer	1.712 (1.730)	1.981 (1.873)	1.962 (1.108)	52

Note: In Game 3, the raw data for calculating the mean and standard deviation for the trustee's transfer were obtained by averaging the individual transfers elicited by the strategy method.

Hypothesis 1. If trustees' preferences present concerns for the vulnerability of trust, they will transfer more tokens in the I-treatment than in the N-treatment.

In the I-treatment, higher transfers of tokens by trustees do not imply per se a higher share of trustworthy choices and, consequently, of trustworthy players. In other words, when vulnerability is salient, it is conceivable that trustees' behavioural strategies could imply more generous but not yet fair transfers, which would leave unchanged the share of trustworthy players. Hence, a second hypothesis follows:

Hypothesis 2. If trustees' preferences show concerns for the vulnerability of trust, the share of trustworthy trustees is larger in the I-treatment ("trustworthy" in the sense defined above).

Irrespective of how trust vulnerability affects the level of return transfers and the share of trustworthy people, its transformative capacity may also affect the reciprocity attitudes of trustees. In fact, both an increase in the return transfer levels and a higher share of trustworthy people may result simply from an upward shift of trustees' behavioural strategies. By contrast, a change in the trustees' reciprocity attitudes would require a different association between trustors' and trustees' transfers, or in other words, a change in the slope of trustees' *return function*, the amount returned as a function of the amount received.

Hypothesis 3. Assuming concerns for the vulnerability of trust, the slope of untrustworthy trustees' return function is higher in the I-treatment.

5 Results

Table 1 reports summary statistics by treatment, game and type of players. First of all, both trustors and trustees transfer on average non-zero amounts of tokens to their co-player. In Game 1 of the N-treatment (vs. I-treatment), trustors sent on average 3.077 (vs. 2.846) tokens to their respective co-players; trustees responded by sending back on average 1.885 (vs. 1.712), which are still positive but lower than what full reciprocity would imply. A set of t-tests and Wilcoxon rank-sum tests (p-values > 0.40) demonstrate that there is no statistically significant difference in Game 1 between the average amounts sent by trustor (vs. trustee) across treatments, showing a successful random assignment of subjects in treatments and roles.

Game 2 of the N-treatment was a close replica of the previous game outcomes. On average, trustors transferred 3.192 tokens to their co-players who responded by sending back 1.885 tokens. In the I-treatment a slight change in the average behaviour was reported in Game 2: trustors and trustees transferred 2.365 and 1.981 tokens, respectively.

Game 3 presents a different overall picture. In the N-treatment (vs. I-treatment), trustors transferred on average 2.135 (vs. 2.731) tokens to trustees who in turn sent back 1.474 (vs. 1.962) tokens.

So far it is worth noticing that Game 1 was designed to identify trustees' trustworthiness; Game 2 was designed to provide subjects with a first experience of the environment we aim to study, while adopting a more intuitive choice mode with a single unconditional decision.

By eliciting the entire set of conditional decisions, Game 3 was designed to efficiently provide the relevant information to fully test the hypotheses stated in the previous section, concerning the effect of vulnerability. Hence, following the approach by Fischbacher et al. (2001), we now focus our analysis largely on the set of trustees' conditional decisions from Game 3.[5] Game 3, in using the strategy method, required

[5]Game 2 did not provide sufficient data for a test of the effect of vulnerability. The critical cases are those in which the trustee was not trustworthy in Game 1, yet, despite this, the trustor transferred 4 or 5 tokens. There were only 20 such cases, 13 in the N-treatment and 7 in the I-treatment. When the trustor transferred less than 4, as we shall see, information had

https://doi.org/10.1017/S1930297500006276 Published online by Cambridge University Press

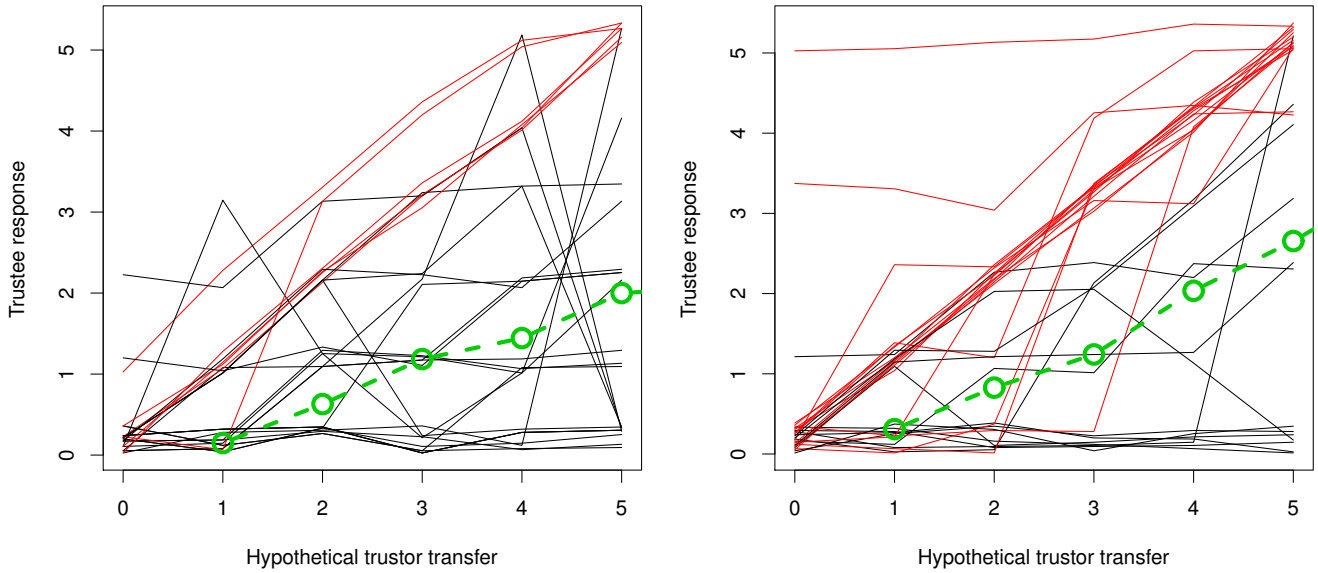

Figure 1: Tokens sent back in the N-treatment (left panel) and I-treatment (right panel) from Game 3 (strategy method).

trustees to indicate how they would respond *if* the trustor were vulnerable in this test. Thus, every trustor provided a response to this hypothetical case.

Figure 1 shows the individual Game 3 return functions (amount returned as a function of the amount transferred by the trustor) for the trustees who were not trustworthy in Game 1, separately for the N-treatment and I-treatment. It is apparent that the mean trustee return function (dashed lines, with circular points) was a roughly linear function of the amount transferred up to about 3 tokens, at which point the trustee response start to become sensitive to the information condition. In particular, trustees returned more when they knew that the trustor was vulnerable, that is, the trustor transferred 4 or 5 tokens despite knowing that the trustee was not trustworthy (Hypothesis 1).

In Figure 1, red lines indicate that the intercept of the linear model for each trustee was greater than 4 (as explained shortly). It is apparent that many more trustees reciprocated in full (i.e., in a way that would count as "trustworthy") in the I-treatment than in the N-treatment (Hypothesis 2, for transfers of 4 or 5, $p = .026$, one-tailed Wilcoxon test on the number of trustworthy responses [0, 1, or 2]).

To test the main hypotheses of interest (Hypotheses 1 and 3), we fit a straight line to the each trustee's return function. Two parameters of this fit were of interest: the intercept when the trustor transferred 5 tokens, and the slope of the line. The intercept assesses the response when the trustor was most vulnerable. The slope assesses the sensitivity of the trustor's response to the amount transferred by the trustor. Of particular interest were the responses of the 56 trustees who were not trustworthy in Game 1.

T tests confirmed the apparent results of Figure 2. The mean intercept was higher in the I-treatment (3.29) than the N-treatment (2.25, $t = 1.97$, $p = .027$ one tailed), and the slope was also higher (.62 vs. .40, $t = 1.88$, $p = .032$ one tailed). The fact that the slope was higher casts doubt on an interpretation in terms of increased altruism alone: the result depends on the vulnerability resulting from a large transfer from the trustor. It represents an effect of vulnerability on reciprocity.

However, we found that both intercept and slope were also strongly affected by individual differences in the trustees' general willingness to return, as measured by their amount returned in Game 2. (Game 1 did not provide additional information; in regression models, its contribution was not significant once Game 2 was included.) The amount returned in Game 2 correlated .70 with the intercept and .52 with the slope, within the trustees who were untrustworthy in Game 1. The Game 2 returns were thus nuisance variables, which contributed extraneous variance to the t tests just reported. They represented pre-existing individual differences in trustworthiness.

To reduce the effect of this extraneous variance, we regressed slope and intercept on information treatment and Game 2 returns, for the trustees who were not trustworthy in Game 1. The regression coefficient for the effect of information treatment on intercept was 1.16 ($p = .001$ one tailed), and the coefficient for the effect on slope was 0.24 ($p = .010$ one tailed). For trustees who were trustworthy in Game 1, information condition had no effect on slope or intercept.

In sum, the results support the hypothesis that previously untrustworthy trustees are more likely to reciprocate high transfers when they know that the trustor knows that they were previously untrustworthy than when they do not know.

little effect, as the trustor was not so obviously vulnerable.

https://doi.org/10.1017/S1930297500006276 Published online by Cambridge University Press

6 Concluding remarks

Our experiment has consistently shown across treatments that when trustors' vulnerability is linked to the trustee's revealed trustworthiness and is made salient by providing relevant information to the players, trustees do change their behaviour by increasing the amount of tokens transferred. In those circumstances, both trustworthy and untrustworthy trustees make more generous transfers, and the degrees of trustworthiness and reciprocity of trustees' behaviour rise. Thus, the transformative nature of vulnerable trust finds consistent support as shown by its capacity of generating higher, more trustworthy and reciprocal transfers by trustees. Vulnerability has shown a transformative capacity.

Our empirical regularities may serve as additional "exhibits" (Sugden, 2003) to be viewed – from a theoretical standpoint – through the lens of intention-based theories, which model other-regarding preferences in the form of reciprocity towards co-players (Rabin, 1993; Dufwenberg & Kirchsteiger, 2004) or aversion to guilt resulting from unfulfilled expectations (Battigalli & Dufwenberg, 2007).[6]

Furthermore, it is easy to envision relevant fields where our results may suggest policy implications and, in general, reflections and suggestions. For instance, a manager adopting subsidiarity management should intervene in team decisions only for activities that would be worse without her *subsidiary* intervention (Melé, 2004). For subsidiarity management to work, it is essential that team members experience managers' genuine, vulnerable trust. Manager should then avoid trying to control or "contractualise" the entire process to prevent possible abuse of trust. A key issue in subsidiarity management is the resilience after a crisis due to betrayal of trust when untrustworthiness is known to the organization, which wants to keep its culture of trust. Our results support the effectiveness of subsidiarity and the importance of giving new trust to team members who appeared to be untrustworthy.

Subsidiarity is essential also in education, where teachers have to create an environment of genuine trust in order to elicit responsibility and freedom. Trusting children, youngsters, and adults with a past of untrustworthiness is a key factor on which the success of the education process hinges. Our results suggest that making salient trustors' (i.e., teacher or social worker) vulnerability may produce a truly transformative effect on trustees (Horsburgh, 1960).

Finally, we hope that our study will stimulate replications and further research to accumulate systematic knowledge on trust in non-enforceable, personalised interactions, and may promote trust as behavioural disposition, as social norms (e.g., Baron, 1998; Dunning et al., 2014), in organisations and beyond.

[6]Intention-based models of aversion to guilt may encompass complementary evidence as, for instance, the one reported by Butler et al. (2016), who link individuals' cheating notions to guilt aversion, and trust.

References

Baier, A. (1986). Trust and Antritrust. *Ethics*, *96*(2), 231–260.

Balliet, D., & Van Lange, P. A. M. (2013). Trust, conflict, and cooperation: A meta-analysis. *Psychological Bulletin*, *139*(5), 1090–1112.

Baron, J. (1998). Trust: Beliefs and morality. In A. Ben-Ner & L. Putterman (Eds.), *Economics, values, and organization* (pp. 408–414). Cambridge: Cambridge University Press.

Battigalli, P., & Dufwenberg, M. (2007). Guilt in games. *American Economic Review*, *97*(2), 170–176.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122–142.

Blount, S. (1995). When social outcomes aren't fair: the effect of causal attributions on preferences. *Organizational Behavior and Human Decision Processes*, *63*(2), 131–144.

Bohnet, I. (2008). Trust in experiments. In Durlauf S.N. & Blume L.E. (Eds.), *The new Palgrave dictionary of economics* (2nd edition). London: Palgrave Macmillan. Retrieved from http://www.dictionaryofeconomics.com/article?id=pde2008_T000241.

Bruni, L. (2012). *The wound and the blessing. Economics, interpersonal relations, happiness*. New York, NY: New City Press.

Butler, J., Giuliano, P., & Guiso, L. (2016). Trust and cheating. *Economic Journal, 126*(595), 1703–1738.

Cialdini, R. B. (2009). *Influence: Science and practice* (5th edition). Boston, MA: Pearson.

Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior*, *47*(2), 268–298.

Dunning, D., Anderson, J. E., Schlösser, T., Ehlebracht, D., & Fetchenhauer, D. (2014). Trust at zero acquaintance: More a matter of respect than expectation of reward. *Journal of Personality and Social Psychology, 107*(1), 122–141

Falk, A., Fehr, E., & Fischbacher, U., (2008). Testing theories of fairness—intentions matter. *Games and Economic Behavior*, *62*(1), 287–303

Fehr, E. (2009). On the economics and biology of trust. *Journal of the European Economic Association*, *7*(2-3), 235–266.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental economics, 10*(2), 171–178.

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters, 71*(3), 397–404.

Holton, R. (1994). Deciding to trust, coming to believe. *Australian Journal of Philosophy*, *72*(1), 63–76.

https://doi.org/10.1017/S1930297500006276 Published online by Cambridge University Press

Horsburgh, H. J. N. (1960). The ethics of trust. *The Philosophical Quarterly*, *10*(41), 343–354.

Johnson, N. D., and Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, *32*(5), 865–889.

Malhotra, D. (2004). Trust and reciprocity decision: The differing perspectives of trustors and trusted parties. *Organizational Behavior and Human Decision Processes, 94*(2), 61–73.

Melé, D. (2004). Exploring the principle of subsidiarity in organizational forms. *Journal of Business Ethics*, *60*(3), 293–305.

Miller, D. T. (1999). The norm of self-interest. *American Psychologist, 54*(12), 1053–1060.

Nussbaum, M. (1986). *The fragility of goodness*. Cambridge: Cambridge University Press.

Pelligra, V. (2007). The not-so-fragile fragility of goodness: The responsive quality of fiduciary relationships. In P.L. Porta & L. Bruni (Eds.), *Handbook of Happiness in Economics*. Cheltenham: Edward Elgar.

Pillutla, M., Malhotra, D., & Murnighan, J. K. (2003). Attributions of trust and the calculus of reciprocity. *Journal of Experimental Social Psychology*, *39*(5), 448–455.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review*, *83*(5), 1281–1302.

Rand, D. G. (2016). Cooperation, fast and slow: Meta-analytic evidence for a theory of social heuristics and self-interested deliberation. *Psychological Science*, *27*(9), 1192–1206.

Ratner, R. K., & Miller, D. T. (2001). The norm of self-interest and its effects on social action. *Journal of Personality and Social Psychology*, *81*(1), 5–16.

Regan, R. T. (1971). Effects of a favor and liking on compliance. *Journal of Experimental Social Psychology*, *7*(6), 627–639.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: a cross-discipline view of trust. *Academy of Management Review*, *23*(3), 393–404.

Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An integrative model of organizational trust: Past, present, and future. *Academy of Management Review, 32*(2), 344–354.

Sugden, R., (2005). Experiments as exhibits and experiments as tests. *Journal of Economic Methodology*, *12*(2), 391–302.

Stanca, L., Bruni, L., & Corazzini, L. (2009). Testing theories of reciprocity: Do motivations matter? *Journal of Economic Behavior & Organization, 71*(2), 233–245.

Stanca, L. (2010). How to be kind? Outcomes versus intentions as determinant of fairness. *Economic Letters*, *106*(1), 19–21.

Strassmair, C. (2009). Can intention spoil the kindness of a gift? An experimental study. Munich Discussion paper No. 2009-4. Retrieved from: https://epub.ub.uni-muenchen.de/10351/1/intentions_march19.pdf.