Judgment and Decision Making, Vol. 11, No. 5, September 2016, pp. 527–536

# Anonymity and incentives: An investigation of techniques to reduce socially desirable responding in the Trust Game

Isabel Thielmann* Daniel W. Heck† Benjamin E. Hilbig ‡§

**Abstract**

Economic games offer a convenient approach for the study of prosocial behavior. As an advantage, they allow for straightforward implementation of different techniques to reduce socially desirable responding. We investigated the effectiveness of the most prominent of these techniques, namely providing behavior-contingent incentives and maximizing anonymity in three versions of the Trust Game: (i) a hypothetical version without monetary incentives and with a typical level of anonymity, (ii) an incentivized version with monetary incentives and the same (typical) level of anonymity, and (iii) an indirect questioning version without incentives but with a maximum level of anonymity, rendering responses inconclusive due to adding random noise via the Randomized Response Technique. Results from a large ($N$ = 1,267) and heterogeneous sample showed comparable levels of trust for the hypothetical and incentivized versions using direct questioning. However, levels of trust decreased when maximizing the inconclusiveness of responses through indirect questioning. This implies that levels of trust might be particularly sensitive to changes in individuals' anonymity but not necessarily to monetary incentives.

Keywords: trust game; social desirability; incentives; anonymity; randomized response technique.

## 1 Introduction

Grounded in game theory, economic games have become a well-established approach across scientific disciplines to study various aspects of (pro)social behavior such as trust and cooperation. A key advantage of economic games is that they model real-life conflicts between certain motives or intentions — for example, individual versus collective utility maximization — in a distilled way. In consequence, economic games allow going beyond mere self-reports of prosocial tendencies in facilitating the measurement or, more specifically, observation of *actual behavior* (Baumeister, Vohs & Funder, 2007). This feature is of particular importance because prosocial behavior is per se socially desirable and self-reports of prosocial behavior are, in turn, prone to over-reporting (e.g., Balcetis & Dunning, 2008; Epley & Dunning, 2000). Specifically, self-report prosociality scales might not only foster biased responding due to socially desirable item content but might also motivate individuals to establish and/or maintain a positive (prosocial) reputation. Economic games, in turn, overcome this limitation by enabling straightforward implementation of techniques to reduce potentially distorting influences of social desirability and to elicit individuals' true behavioral preferences. Most prominently, these are incentives and anonymity.

Incentives constitute a standard in economic research (Camerer & Hogarth, 1999; Hertwig & Ortmann, 2001). In essence, incentives denote behavior-contingent payoffs that individuals receive based on their own and (depending on the game structure) others' behavior in the game. That is, unlike in self-reports, hypothetical scenarios, and the like, individuals actually interact with one or more real others, and all receive (typically monetary) payoffs according to their decisions in the game. In many cases, incentives are simply given as windfalls by the experimenter, without requiring the individual to exert any effort to earn her endowment. In the current work, we will focus on such windfall incentives. Other prior research has also relied on earned incentives, thus emphasizing the role of property rights and asset legitimacy for behavior in economic games (e.g., Cherry, Frykblom & Shogren, 2002; Hoffman, McCabe, Shachat & Smith, 1994; List, 2007). In general, incentives are assumed to constitute one way to reduce self-presentational concerns because behavior is truly consequential, thus motivating individuals to behave as they would actually behave in equivalent real-life situations — usually more selfishly and less prosocially. Correspondingly, it has been noted that "incentives are … useful when the response may be affected by social desirability, as in the case of cooperation in social dilemmas" (Baron, 2001, p. 403; for similar reasoning see, e.g., Camerer & Hogarth, 1999; Gneezy, Meier & Rey-Biel, 2011).

The work reported herein was supported by a grant to the first author from the research focus "Communication, Media, and Politics" funded by the research initiative of Rhineland-Palatinate, Germany



*University of Koblenz-Landau, Department of Psychology, Fortstraße 7, 76829 Landau, Germany. Email: thielmann@uni-landau.de.

†University of Mannheim.

‡University of Koblenz-Landau.

§Max Planck Institute for Research on Collective Goods.



https://doi.org/10.1017/S1930297500004605 Published online by Cambridge University Press

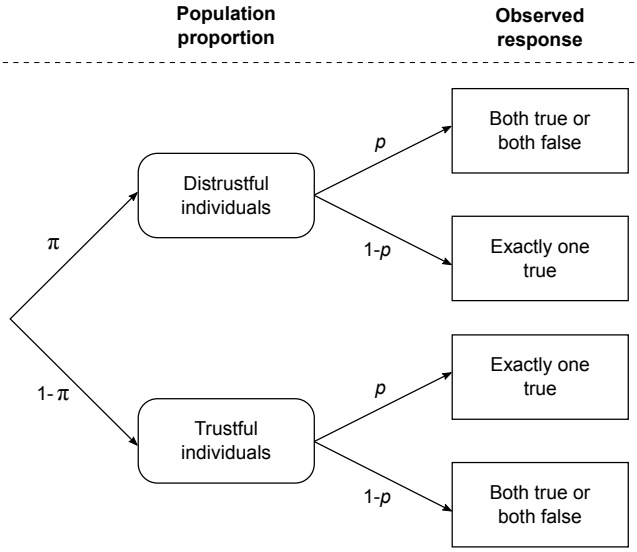
However, evidence on this conjecture is mixed, at least when behavior in economic games with windfall incentives is the dependent measure. (For reviews on the effectiveness of incentives more generally, see, e.g., Camerer & Hogarth, 1999; Gneezy et al., 2011; Hertwig & Ortmann, 2001; for recent evidence, see DellaVigna & Pope, 2016.) On the one hand, some studies indeed suggest that individuals are less generous in incentivized compared to hypothetical settings — as, for example, shown in the Prisoner's Dilemma Game, the Dictator Game (and variants thereof), and the Ultimatum Game (Bühren & Kundt, 2015; Camerer & Hogarth, 1999; Fantino, Gaitan, Kennelly & Stolarz-Fantino, 2007; Gillis & Hettler, 2007; Lönnqvist, Verkasalo & Walkowitz, 2011) — and that the influence of individuals' personality on prosocial behavior differs as a function of whether incentives are provided or not (Balliet, Parks & Joireman, 2009; Ben-Ner, Kramer & Levy, 2008; Lönnqvist et al., 2011). On the other hand, other evidence shows equivalent or even increased levels of prosocial behavior in incentivized compared to hypothetical scenarios. For example, findings from the Trust Game suggest that levels of trust increase when real rather than hypothetical incentives are at stake (Fetchenhauer & Dunning, 2009; Holm & Nystedt, 2008). Furthermore, a broad meta-analysis including more than 100 studies on the Dictator Game revealed that the amount dictators were willing to give were highly similar for real versus hypothetical (windfall) money and interactions (Engel, 2011; for a re-analysis of findings, see also Zhang & Ortmann, 2014). Comparable null effects of incentives have been observed in other studies using variants of the Dictator Game (Ben-Ner et al., 2008; Ben-Ner, McCall, Stephane & Wang, 2009; Brañas-Garza, 2006), the Ultimatum Game (Cameron, 1999), and the Public Goods Game (Gillis & Hettler, 2007). Overall, then, findings on the impact of incentives and thus, by implication, their potential effectiveness in reducing socially desirable responding are inconclusive.

However, a potential reason for this apparent inconclusiveness of findings regarding behavior in economic games might be that the implementation of behavior-contingent incentives necessarily decreases anonymity because an individual's behavior is typically revealed to the experimenter or other participants based on the payoffs she receives (Zizzo, 2010). This decreased anonymity (in the sense that responses are *visible* to others) might, in turn, increase self-presentational concerns, thus outweighing the power of incentives to reduce socially desirable responding. To counteract this drawback, anonymous payment strategies (e.g., double-blind protocols) can be used to maintain invisibility of responses, even if incentives are provided. Thereby, selfish behavior is no longer incriminating, simply because it cannot be inferred from an individual's payoffs. For example, in the double-blind Dictator Game (Hoffman et al., 1994), individuals receive an envelope containing a pre-specified monetary endowment, decide in private how much of this endowment to keep for themselves, and finally place the envelope with the remaining money for the recipient into a hidden ballot box (e.g., Eckel & Grossman, 1996; Hilbig, Thielmann, Hepp, Klein & Zettler, 2015). By this means, an individual's behavior is perfectly concealed because neither the experimenter nor other participants can observe the individual payoff. Similarly, in web-based studies, anonymity can be ensured despite behavior-contingent payment if incentives are paid out by an external panel provider who is unfamiliar with the specific payoff scheme. As such, an individual's behavior can again neither be inferred by the experimenter, nor by other participants or the panel provider (e.g., Hilbig & Zettler, 2015; Thielmann, Hilbig, Zettler & Moshagen, in press). In general, such approaches will ensure that incentivized and hypothetical scenarios are comparable regarding the level of anonymity they warrant, namely invisibility of responses.

As mentioned above, anonymity *per se* is an efficient approach to reduce self-presentational concerns and social desirability, respectively. One way to increase anonymity still further is through techniques that render responses not only invisible but also inconclusive by adding random noise to participants' responses. In consequence, responses no longer mirror an individual's actual behavior. Such an approach of rendering responses inconclusive is, for example, realized in indirect questioning techniques and, more specifically, the Randomized Response Technique (RRT; Warner, 1965) which "guarantees privacy and may well overcome respondents' reluctance to reveal sensitive … information" (Lensvelt-Mulders, Hox, Van Der Heijden & Maas, 2005, p. 321). For example, in the Crosswise Model (CWM; Yu, Tian & Tang, 2007; see Figure 1) — a revised version of the original RRT model proposed by Warner (1965) on which we will rely in the present study — participants are simultaneously presented with two statements and simply asked to indicate whether (i) both statements are true or both statements are false or (ii) only one of the two statements is true. In particular, one statement refers to a critical (i.e., sensitive) issue under scrutiny which is potentially prone to socially desirable responding, such as prosocial versus selfish behavior in an economic game. The prevalence of this *sensitive attribute* is, by definition, unknown to the experimenter. The other statement, in contrast, refers to an independent, non-sensitive issue for which socially desirable responding is unlikely to occur, such as a participant's month of birth. Importantly, so long as a participant's month of birth is unknown to the experimenter, it is impossible to unequivocally infer her response to the critical (sensitive) question (i.e., prosocial vs. selfish behavior) because both the sensitive and non-sensitive statements are answered in combination. Anonymity is hence maximized by rendering responses inconclusive. However, given that the prevalence of the randomization device is known by design (e.g., partic-

https://doi.org/10.1017/S1930297500004605 Published online by Cambridge University Press

Figure 1: Tree diagram of the Crosswise Model (CWM; Yu et al., 2008) as implemented in the RRT version of the Trust Game.




ipants' month of birth is known from official birth statistics), the level of selfishness can be estimated at the group level.

Of critical importance for the issue at hand, the RRT approach has already been successfully implemented in economic games. Specifically, in a hypothetical Prisoner's Dilemma Game, cooperation rates have been shown to substantially decrease when socially desirable responding was prevented by RRT instructions (Moshagen, Hilbig & Musch, 2011). Similarly, in the Dictator Game, giving decreased when rendering responses inconclusive via RRT compared to implementing a double-blind protocol (Franzen & Pointner, 2012). Hence, although research relying on indirect questioning techniques in economic games is currently scarce, previous evidence implies that the RRT is a feasible and effective method to reduce socially desirable responding in this context. (For a meta-analysis supporting the validity of the RRT *in general*, see Lensvelt-Mulders et al., 2005.)

Summing up, the current state of research implies that behavior-contingent incentives as well as anonymity (in terms of invisibility and inconclusiveness of responses) might be promising approaches to reduce self-presentational concerns and thus influences of social desirability on prosocial behavior in economic games. However, corresponding evidence is available only for some prosocial behaviors and game paradigms (such as giving in the Dictator Game) but not for others. The present study aimed at closing this gap for another heavily studied prosocial behavior: trust in the Trust Game (Berg, Dickhaut & McCabe, 1995). Despite the abundant use of the Trust Game (see Thielmann & Hilbig, 2015, for a recent review), research on whether incentives are decisive for trust behavior is scarce, with only one study directly comparing behavior following a random assignment to an incentivized game (albeit without implementing an anonymous payment strategy) versus a hypothetical game (Fetchenhauer & Dunning, 2009). More importantly, techniques rendering responses inconclusive have never been applied to the Trust Game.

Table 1: Overview of Trust Game versions used.

| Game version | Questioning type | Anonymity of responses | Incentives |
|---|---|---|---|
| Hypothetical | Direct | Invisibility | No |
| Incentivized | Direct | Invisibility | Yes |
| RRT | Indirect | Inconclusiveness | No |

Note. RRT = Randomized Response Technique.

In the present study, we used three versions of the Trust Game to demonstrate the applicability of said techniques to this specific game paradigm and to test their potential effects on trust behavior: a direct questioning (DQ) hypothetical version without behavior-contingent incentives or real interaction but with invisibility of responses, a DQ incentivized version with behavior-contingent incentives, real interaction, and invisibility of responses, and an indirect questioning version following the RRT approach, again without behavior-contingent incentives or real interaction, but with maximum anonymity due to inconclusiveness of responses (see Table 1 for an overview). To implement the RRT, we relied on the CWM given that this specific RRT design is characterized by particularly simple instructions and has also been shown to be superior to other RRT models (Hoffmann & Musch, 2015). In general, we considered it important to rely on a large and representative sample to ensure that our conclusions are as generalizable as possible to applications of the Trust Game in diverse samples.

# 2 Method

## 2.1 The Trust Game

In the classical Trust Game (Berg et al., 1995), a trustor and a trustee both receive an initial endowment. The trustor, on whom we focus in what follows, is first asked how much of her endowment she wants to transfer to the trustee, thus measuring her level of trust. The transferred amount is multiplied (typically tripled) by the investigator and added to the trustee's endowment who, in return, decides how much to transfer back to the trustee, thus measuring her level of trustworthiness. Hence, given that trust maximizes social welfare (i.e., the trustor's transfer is multiplied) and also ensures that the trustee has any choice to make, trust arguably constitutes the prosocial and thus socially desirable choice

https://doi.org/10.1017/S1930297500004605 Published online by Cambridge University Press

(for similar reasoning see Fehr, Fischbacher, von Rosenbladt, Schupp & Wagner, 2003; Naef & Schupp, 2009).

In the present study, we used a binary version of the Trust Game, which all participants completed as trustors (and some participants additionally as trustees; see below). Trustors, as well as trustees, were initially endowed with 3€ (approximately US$3.20 at the time of data collection). Meta-analytic evidence (Johnson & Mislin, 2011) has revealed that endowments of this size suffice to make the game "real" and that, in general, trust behavior does not seem to vary as a function of incentive size. Participants (trustors) were asked to decide whether or not they want to transfer their 3€ endowment to the trustee (who was simply called "the other"). If a participant decided to transfer the amount, it was tripled to 9€ and added to the trustee's endowment — who could, in turn, decide how much of the 9€ to return to the trustor. Thus, each dyad earned either 6€ (in case of distrust; split exactly 50:50) or 12€ (in case of trust; split depending on the trustee's decision).

As sketched above, participants were randomly assigned to one of three versions of the Trust Game as trustors: the hypothetical, the incentivized, or the RRT version (Table 1). In the hypothetical (DQ) version, participants were simply asked to imagine interacting with another unknown person for real money. That is, participants were fully aware about the hypothetical nature of the game and provided their trust decision of whether to transfer their hypothetical 3€ to the other "as if" the situation was real. Thus, individuals' anonymity was preserved in that responses were only related to an anonymous ID (as is common practice) and were hence basically invisible, meaning that an individual's behavior was neither revealed to the experimenter, nor to the panel provider or other participants.

In the incentivized (DQ) version, participants made the same trust decision as in the hypothetical version but, in contrast, knew that they would be paid according to their own and a trustee's behavior. Thus, participants interacted with a real other for real money. Note that, to implement this real interaction, participants assigned to the hypothetical version also acted in the role of the trustee in the incentivized game (following their response as a trustor, thus ensuring that trust behavior — as focused on herein — was unaffected by this procedure). However, these data were entirely omitted in the analyses. In general, to ensure that individuals' responses remained anonymous — in terms of being invisible to others (including the experimenter) — payment was handled by the panel provider, who was entirely unfamiliar with the payment scheme of the present study. This ensured similar levels of anonymity in the two DQ versions (Table 1).

Finally, in the RRT version, we again implemented a hypothetical design given that the procedure, by implication, forbids inferring individuals' actual decision in the game, thus rendering the payment of truly behavior-contingent incentives impossible. That is, similar to the hypothetical version, participants were asked to imagine interacting with another unknown person for real money. Participants further received the information that a statistical procedure would be used which guarantees perfect anonymity. Specifically, participants were simultaneously presented with two statements A and B (following the CWM design), namely

- Statement A: "I keep the 3.00€ for myself and do not transfer anything to Player 2."
- Statement B: "I was born in April or May."

and asked whether they agree (i) to both or none of the statements or (ii) to exactly one statement (but not the other). Notably, due to these simple instructions, the CWM features high comprehensibility and compliance of participants and thus ensures valid prevalence estimates (Hoffmann & Musch, 2015). Using participants' month of birth as a randomization device allowed to further keep the randomization procedure as simple as possible (Moshagen et al., 2011). Importantly, the corresponding probabilities are known by design (i.e., $p = 2/12$ for being born in April or May in Germany, corresponding to official statistics; Statistisches Bundesamt, 2012)[1] whereas it is clear to participants that their individual month of birth is unknown to the investigator. Thus, responses are maximally anonymous in the sense of being inconclusive regarding participants' behavior. Notably, this was the main difference between the hypothetical DQ and the RRT game version in the current study (Table 1).

## 2.2 Procedure

The study was run via the Internet, closely adhering to common guidelines for web-based experimenting (Reips, 2002a, 2002b). After providing consent, participants provided demographic information and received detailed instructions on the binary Trust Game, in the role of the trustor for one of the three Trust Game versions (hypothetical, incentivized, or RRT). All participants then provided their trust responses of whether to transfer their 3€ to the trustee or not (in combination with one's month of birth in the RRT game version). Participants assigned to the hypothetical version additionally (and subsequently) indicated how much of the tripled trust transfer of 9€ they would return to the trustor as a trustee in case the trustor actually decides to invest her 3€. After completing data collection, participants assigned to the incentivized version as either trustor or trustee were randomly matched and paid according to their own and the other's behavior in the game ($M = 4.62$€, $SD = 2.73$€) by the panel provider. In addition to the behavior-contingent incentives, all participants received a flat fee of 2€.

[1] Of course, the probability of the randomization device may slightly differ in the sampled population. Importantly, however, the estimated trust rates in our data remained virtually the same when considering slightly different probabilities than $p = 2/12$, as summarized in the supplementary analyses.

https://doi.org/10.1017/S1930297500004605 Published online by Cambridge University Press

## 2.3 Participants

Given that RRT designs are typically analyzed in a multinomial processing tree framework, we used the free software multiTree (Moshagen, 2010) for an a priori power analysis. Based on the expected effect size, the significance level, and the desired power, multiTree uses an iterative algorithm to find the correspondingly required sample size. For the effect sizes, we assumed identical trust rates of 50% in the hypothetical and the incentivized game — based on meta-analytic average trust levels (Johnson & Mislin, 2011) and the currently inconclusive evidence on whether incentives are influential. In contrast, we expected a slightly lower trust rate of 35% in the RRT condition — based on prior findings showing reduced prosocial behavior when using the RRT in economic games (Franzen & Pointner, 2012; Moshagen et al., 2011). Moreover, we aimed at a high power of $1 - \beta$ = .95 to detect a different trust rate in the RRT version compared to any of the two DQ versions of the game (i.e., hypothetical and incentivized) with $\alpha$ = .05. Overall, this analysis resulted in a required sample size of $N$ = 1,228 participants. As is common practice in RRT research, we counteracted the increased error variance inherent in the RRT design due to adding random noise (e.g., Moshagen, Musch & Erdfelder, 2012; Moshagen & Musch, 2012) by assigning twice as many participants to the RRT version of the game ($n$ = 614) than to the hypothetical and incentivized DQ versions ($n$ = 307 each). Correspondingly, we recruited $N$ = 1,267 participants who completed the study and were thus included in the data analysis (out of $N$ = 1,367 starting the study, i.e. 92.7% completion rate.[2]) Specifically, $n$ = 324 completed the Trust Game in the hypothetical version, $n$ = 317 in the incentivized version, and $n$ = 626 in the RRT version.

As mentioned above, recruitment of participants was carried out by a professional panel provider (i.e., the German company *respondi AG*), thus allowing for collecting a diverse and heterogeneous (non-student) sample (Henrich, Heine & Norenzayan, 2010). In particular, participants were almost equally distributed across the sexes (43.2% female) and also covered a broad age range (from 18 to 65 years), with a higher average age (and standard deviation) than is typically observed for student samples ($M$ = 41.1, $SD$ = 12.5 years). Moreover, there was a substantial diversity in educational level, with 36.5% holding a general certificate of secondary education (German: Realschulabschluss), 29.4% a vocational diploma or university-entrance diploma (German: Fachabitur or Abitur), and 31.6% a university/college degree. The majority of participants (70.2%) were employed; only 7.7% were students. A more detailed overview of the sample composition in each game version is available online in the analysis supplement. The same applies to analyses taking into account the demographic data (i.e., sex and age) as control variables — which yielded similar results.

[2]The largest dropout occurred in the RRT version of the game (8.9%), followed by the incentivized version (6.8%) and the hypothetical version (4.7%).

## 2.4 Statistical analyses

Analyses were based on the trust rates observed in each version of the Trust Game. In the hypothetical and incentivized DQ versions, classical frequentist estimates for the trust rates are directly available as the respective proportions of trustors entrusting their 3€. In the RRT version, however, the trust rate cannot be directly observed, but only the proportion $\lambda$ of participants stating that either both statements are true or both statements are false (with one statement referring to their decision to distrust and the other to their month of birth being April or May). This observed proportion $\lambda$ is hence determined by the underlying prevalence of distrust $\pi$ (or equivalently, the trust rate $\tau = 1 - \pi$) and the probability $p$ that a participant's month of birth is April or May ($p$ = 2/12), as follows (see Figure 1):

$$\lambda = \pi p + (1 - \pi)(1 - p) \tag{1}$$

Solving Equation 1 for the trust rate $\tau = 1 - \pi$ results in the maximum-likelihood estimator[3] (Yu et al., 2007):

$$\hat{\tau} = \frac{p - \lambda}{2p - 1}. \tag{2}$$

Thus, given that the corresponding probability $p$ of the randomization device (i.e., month of birth) is known, straightforward calculations permit estimating the trust rate $\tau$ in the sample (for details on standard errors and confidence intervals for this estimate see Yu et al., 2007).

To further compare the trust rates across the three game versions, likelihood ratio tests allow comparing the full model specifying separate trust rates for each game version with a nested model constraining trust rates to be identical. We performed these analyses using the R package RRreg, which implements univariate and multivariate analyses for RRT designs (Heck & Moshagen, 2016), and we double-checked the results using multiTree (Moshagen, 2010) — which led to identical results. Note that, in general, such an approach of comparing the trust rates rests on the premise that potential differences in the prevalence of (dis)trust across game versions are readily interpretable in terms of differential effectiveness of the corresponding technique(s) to reduce socially desirable responding (Moshagen, Hilbig, Erdfelder & Moritz, 2014).

[3]Given that distrust refers to the sensitive issue under scrutiny, we relied on the decision to *keep* (rather than transfer) the money in the RRT instructions (see above, Statement A). However, for the sake of consistency with prior research, we herein refer to trust (i.e., whether individuals decided *not* to keep their money) as the dependent measure of interest, thus solving Equation 2 for the trust rate rather than the distrust rate.

https://doi.org/10.1017/S1930297500004605 Published online by Cambridge University Press

Moreover, we complemented the classical frequentist analyses by Bayesian analyses (see Lee & Wagenmakers, 2013, for an introduction). Unlike frequentist analyses, Bayesian analyses take prior beliefs about the parameters into account — in our case, we used uninformative uniform prior distributions on the interval (0,1) for the trust rate in each of the three game versions. Based on these priors (and the same likelihood function as in the frequentist analyses), posterior estimates and credible intervals for the trust rates can be calculated. The software JAGS (Plummer, 2003) provides a useful tool for this purpose which is based on Markov chain Monte Carlo sampling. As an advantage compared to confidence intervals (CIs), the credible intervals are directly interpretable as providing the 95% most plausible values for the trust rates, conditional on the uniform prior and the data (Morey, Hoekstra, Rouder, Lee & Wagenmakers, 2016). Moreover, we relied on the Savage-Dickey method (Wagenmakers, Lodewyckx, Kuriyal & Grasman, 2010) to compute (pairwise) Bayes factors for comparing the resulting trust rates across conditions. The Bayes factor $B_{10}$ quantifies the evidence in favor of different trust rates ($\mathcal{H}_1$: $\tau_1 - \tau_2 \neq 0$)[4] in relation to evidence in favor of identical trust rates ($\mathcal{H}_0$: $\tau_1 - \tau_2 = 0$), thus providing direct information on the plausibility of the alternative versus null hypothesis. Theoretically, $B_{10}$ is defined as the multiplicative factor that is required to update the prior odds of $\mathcal{H}_1$ versus $\mathcal{H}_0$ to the posterior odds in light of the data. Following the Savage-Dickey method, the Bayes factor $B_{10}$ is computed as the ratio of the posterior to the prior density for a difference in trust rates of $\tau_1 - \tau_2 = 0$.[5] Note that for the sake of interpretability, we report the inverse Bayes factor in favor of the null hypothesis (i.e., $B_{01} = 1/B_{10}$) if the Bayes factor implies support for the null hypothesis.

# 3 Results

Given the apparent differences in complexity of trust questions across game versions, we considered it important to first rule out potential differences in motivation between answering the rather simple DQ questions and the more complex RRT question (Lensvelt-Mulders et al., 2005). Therefore, we compared the time participants required for completing the Trust Game in either version. If participants actually answered the trust question in the RRT version seriously, we expected them to require more time for reading the instructions than participants in the DQ versions. Supporting this hypothesis, average response times were larger in the RRT version ($M$ = 171s, $SD$ = 564s) compared to both the hypothetical version ($M$ = 96s, $SD$ = 233s, $t(705) = 8.95$, $p < .001$, $d = 0.19$, $B_{10} > 1000$) and the incentivized version ($M$ = 95s, $SD$ = 375s, $t(753) = 9.30$, $p < .001$, $d = 0.16$, $B_{10} > 1000$).[6] For the two DQ versions, in turn, no differences in response times were apparent, $t(634) = 0.17$, $p = .862$, $d < .01$, $B_{01} = 11.2$.

Figure 2: Estimated proportion of participants deciding to trust (i.e., trust rates according to frequentist maximum-likelihood estimates) in the Trust Game, separated for the three game versions used (error bars show $\pm 1$ *SE*; DQ = direct questioning; RRT = Randomized Response Technique).

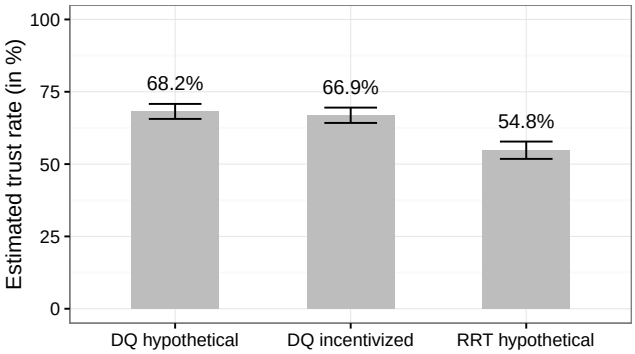



Figure 2 summarizes the estimated trust rates and standard errors for the three versions of the Trust Game used. As is apparent from the left two bars, trust rates were highly similar across the two DQ versions of the game. In particular, in the hypothetical version, 68.2% of trustors decided to entrust their 3€ (95% CI [63.1%, 73.3%]) whereas in the incentivized version, 66.9% of participants did so (95% CI [61.7%, 72.1%]). Finally, analyzing individuals' responses in the (hypothetical) RRT version of the game revealed an estimated trust rate of 54.8% (95% CI [48.9%, 60.7%]; see right bar in Figure 2). In the Bayesian analysis, trust rates were estimated via the means of the posterior samples with 95% credible intervals for the hypothetical (68.1% [63.0%, 73.1%]), the incentivized (66.8% [61.6%, 71.7%]), and the RRT version of the Trust Game (54.8% [48.8%, 60.5%]) Note that these Bayesian estimates closely resembled the maximum-likelihood estimates — which is reasonable given the comparably small impact of the prior compared to the actual data due to our large sample size. However, as sketched above, the credible intervals are — advantageously — directly interpretable as containing the 95% most plausible estimates for the respective trust rates.

Furthermore, we considered it informative to statistically compare these trust rates across game versions using (a) likelihood-ratio tests (with a Bonferroni-corrected signifi-

[4] The assumed prior on the difference $\tau_1 - \tau_2$ is a symmetric triangular distribution from –1 to 1, as implied by independent uniform priors for both trust rates.

[5] Alternatively, marginal likelihoods can be directly computed by numerical integration (Heck, Wagenmakers & Morey, 2015) as illustrated in the online supplementary analyses.

[6] The two-sample $t$-tests were performed on the log-transformed values to account for the right-skewness of response times. Due to inequality of variance across conditions, we relied on the Welsh's $t$-test. Bayes factors, in turn, were calculated using a Bayesian $t$-test with a standard Cauchy prior (Rouder, Speckman, Sun, Morey & Iverson, 2009).

https://doi.org/10.1017/S1930297500004605 Published online by Cambridge University Press

cance level of $\alpha$ = .017 for three pairwise tests) and (b) pairwise Bayes factors. First, comparing the hypothetical and incentivized DQ versions failed to reveal a significant difference in trust rates, $G^2(1) = 0.1$, $p = .719$, despite the high power of our study. Offering a more direct assessment of the relative evidence in favor of $\mathcal{H}_0$ (identical trust rates) versus $\mathcal{H}_1$ (different trust rates), the corresponding Bayes factor of $B_{01} = 10.2$ indicated approximately 10 times more evidence in favor of identical rather than different trust rates in the two DQ versions. Comparing the two hypothetical (i.e., DQ and RRT) versions further revealed a significantly higher trust rate in the hypothetical DQ version than in the RRT version, $G^2(1) = 11.3$, $p < .001$. This suggests that the increased anonymity due to inconclusiveness of responses in the RRT version led participants to trust less. The corresponding Bayes factor further implied substantially more evidence in favor of the alternative than of the null hypothesis, $B_{10} = 27.2$, thus supporting the conclusion of different trust rates. Finally, the RRT version of the game also revealed a lower trust rate than the incentivized version, $G^2(1) = 9.0$, $p = .003$, resulting in a Bayes Factor of $B_{10} = 8.9$. That is, the data yielded almost nine times more support for the alternative hypothesis implying different trust rates than for the null hypothesis implying identical trust rates. However, it should be noted that, by design, the incentivized and the RRT version of the game varied on two factors (i.e., incentives *and* anonymity; see Table 1), thus limiting the direct interpretability of the observed difference in trust rates.

# 4 Discussion

A major advantage of economic games, compared to mere self-reports, is that they facilitate the measurement of actual prosocial behavior and allow for a convenient implementation of strategies to reduce potentially distorting influences of social desirability. The present study aimed to investigate the most prominent of these strategies, namely behavior-contingent incentives and anonymity, in the Trust Game. We examined individuals' willingness to trust in three different game versions: a hypothetical version without real monetary incentives or interaction partner, an incentivized version with real incentives and interaction partner (both preserving anonymity in terms of invisibility of responses), and an indirect questioning version relying on the Randomized Response Technique without monetary incentives or interaction partner, but maximizing anonymity in terms of inconclusiveness of responses. Results were based on a large and heterogeneous (i.e., representative, non-student) sample, thus allowing inferences on diverse applications of the Trust Game in different populations.

Regarding the level of trust in the different game versions, we observed that approximately two thirds of participants opted for trust if directly asked whereas only half of participants did so if their responses were disguised by indirect questioning. Correspondingly, no differences in the willingness to trust an unknown other were apparent between the two direct questioning versions, that is, as a function of whether actual or hypothetical money was at stake (and whether, in turn, the situation involved a "real" other). This implies that hypothetical and incentivized scenarios might indeed yield similar trust rates, at least when the level of anonymity is held constant and does not decrease due to (non-anonymous) behavior-contingent payment. Furthermore, our results suggest that participants take into account whether their responses are truly anonymous or not. That is, comparing the two hypothetical versions (which only differed with regard to the questioning technique and thus the level of anonymity preserved) showed a lower willingness to trust once anonymity was maximized through rendering responses inconclusive via indirect questioning. This implies that social desirability might indeed be a potential driver of trust behavior in the Trust Game — and that maximizing anonymity might be particularly effective in reducing corresponding response biases. More generally, our study demonstrates how different techniques to reduce socially desirable responding can be implemented in the Trust Game. As such, it might build a valuable foundation for future research, both methodologically and with regard to the to-be-expected levels of trust in a representative sample.

Notwithstanding these advantages, however, some limitations should be acknowledged: First, using the RRT as an indirect questioning technique might have produced further differences between game versions than merely maximizing anonymity. For example, the more complex instructions in the RRT version might have triggered more thorough processing of the game situation, thus simply decreasing prosocial behavior by provoking more deliberation (Rand, Greene & Nowak, 2012). Moreover, we cannot be sure that our participants actually understood the RRT instructions given that we did not include a comprehension check. Although the CWM design has generally proven to be highly understandable (e.g., Hoffmann & Musch, 2015) — and our analyses of response times indicated that participants required substantially more time when faced with the RRT version of the Trust Game compared to both DQ versions — future research using the RRT might profit from directly assessing participants' comprehension and compliance (e.g., Hilbig, Moshagen & Zettler, 2015). This might be of particular importance when the RRT is applied to the Trust Game because this specific game paradigm is per se more complex than other paradigms (such as the Dictator Game).

In a similar vein, future studies might consider using different variants of the statements to which participants respond. That is, it might be valuable to investigate whether estimated trust rates remain similar once rephrasing Statement A such that it refers to transferring the money (i.e., trust) rather than keeping it (i.e., distrust). Finally, it should

https://doi.org/10.1017/S1930297500004605 Published online by Cambridge University Press

be noted that anonymity alone might not suffice to undermine socially desirable responding, because individuals might still want to appear non-selfishly to themselves. A viable extension of our design might hence be to additionally reduce situational transparency, thus providing a justification for individuals to behave more selfishly (Dana, Weber & Kuang, 2007).

Importantly, our findings do not strictly imply that incentives are generally ineffective in reducing socially desirable responding in economic games in general or in the Trust Game in particular. That is, although we did not observe any differences in trust rates as a function of incentivization in direct questioning, we cannot rule out that the motives driving behavior differ between hypothetical and incentivized scenarios. For example, if the trustor knows that there is no real trustee (as in the hypothetical version), trust behavior might no longer be based on expectations regarding the other's trustworthiness. Future research is certainly required to clarify this issue.

In conclusion, our findings suggest that trust in the Trust Game might indeed be sensitive to changes in the level of anonymity of individuals' responses. By contrast, providing behavior-contingent incentives does not necessarily alter levels of trust. Although incentives are certainly useful to render the behavior under scrutiny more "real", it should not be taken for granted that incentives will suffice to eliminate socially desirable responding and correspondingly affect levels of trust. This is plausible given that, even with incentives present, people may be influenced by reputational concerns that can arguably be minimized only when responses are inconclusive — as is the case under indirect questioning. Rather than relying on incentives alone to rule out social desirability biases, future research may profit from corresponding alternative methods.

## References


Balcetis, E., & Dunning, D. A. (2008). A mile in moccasins: How situational experience diminishes dispositionism in social inference. *Personality and Social Psychology Bulletin, 34*(1), 102–114. http://dx.doi.org/10.1177/0146167207309201.

Balliet, D., Parks, C. D., & Joireman, J. A. (2009). Social value orientation and cooperation in social dilemmas: A meta-analysis. *Group Processes & Intergroup Relations, 12*(4), 533–547. http://dx.doi.org/10.1177/1368430209105040.

Baron, J. (2001). Purposes and methods [Peer commentary on "Experimental practices in economics: A methodological challenge for psychologists?" by R. Hertwig & A. Ortmann]. *Behavioral and Brain Sciences, 24*(3), 383–451.

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science, 2*(4), 396–403. http://dx.doi.org/10.1111/j.1745-6916.2007.00051.x.

Ben-Ner, A., Kramer, A., & Levy, O. (2008). Economic and hypothetical dictator game experiments: Incentive effects at the individual level. *Journal of Socio-Economics, 37*(5), 1775–1784. http://dx.doi.org/10.1016/j.socec.2007.11.004.

Ben-Ner, A., McCall, B. P., Stephane, M., & Wang, H. (2009). Identity and in-group/out-group differentiation in work and giving behaviors: Experimental evidence. *Journal of Economic Behavior & Organization, 72*(1), 153–170. http://dx.doi.org/10.1016/j.jebo.2009.05.007.

Berg, J., Dickhaut, J., & McCabe, K. A. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10*(1), 122–142. http://dx.doi.org/10.1006/game.1995.1027.

Brañas-Garza, P. (2006). Poverty in dictator games: Awakening solidarity. *Journal of Economic Behavior & Organization, 60*(3), 306–320. http://dx.doi.org/10.1016/j.jebo.2004.10.005.

Bühren, C., & Kundt, T. C. (2015). Imagine being a nice guy: A note on hypothetical vs. incentivized social preferences. *Judgment and Decision Making, 10*(2), 185–190.

Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty, 19*(1-3), 7–42. http://dx.doi.org/10.1023/A:1007850605129.

Cameron, L. A. (1999). Raising the stakes in the ultimatum game: Experimental evidence from Indonesia. *Economic Inquiry, 37*(1), 47–59. http://dx.doi.org/10.1111/j.1465-7295.1999.tb01415.x.

Cherry, T. L., Frykblom, P., & Shogren, J. F. (2002). Hardnose the dictator. *American Economic Review, 92*(4), 1218–1221. http://dx.doi.org/10.1257/00028280260344740.

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory, 33*(1), 67–80. http://dx.doi.org/10.1007/s00199-006-0153-z.

DellaVigna, S., & Pope, D. (2016). *What motivates effort? Evidence and expert forecasts*. NBER Working Paper No. 22193.

Eckel, C. C., & Grossman, P. J. (1996). Altruism in anonymous dictator games. *Games and Economic Behavior, 16*(2), 181–191. http://dx.doi.org/10.1006/game.1996.0081.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics, 14*(4), 583–610. http://dx.doi.org/10.1007/s10683-011-9283-7.

Epley, N., & Dunning, D. A. (2000). Feeling "holier than

thou": Are self-serving assessments produced by errors in self- or social prediction? *Journal of Personality and Social Psychology, 79*(6), 861–875. http://dx.doi.org/10.1037/0022-3514.79.6.861.

Fantino, E., Gaitan, S., Kennelly, A., & Stolarz-Fantino, S. (2007). How reinforcer type affects choice in economic games. *Behavioural Processes, 75*(2), 107–114. http://dx.doi.org/10.1016/j.beproc.2007.02.001.

Fehr, E., Fischbacher, U., von Rosenbladt, B., Schupp, J., & Wagner, G. G. (2003). *A nation-wide laboratory: Examining trust and trustworthiness by integrating behavioral experiments into representative survey*. CESifo Working Paper Series: CESifo Working Paper No. 866.

Fetchenhauer, D., & Dunning, D. A. (2009). Do people trust too much or too little? *Journal of Economic Psychology, 30*(3), 263–276. http://dx.doi.org/10.1016/j.joep.2008.04.006.

Franzen, A., & Pointner, S. (2012). Anonymity in the dictator game revisited. *Journal of Economic Behavior & Organization, 81*(1), 74–81. http://dx.doi.org/10.1016/j.jebo.2011.09.005.

Gillis, M. T., & Hettler, P. L. (2007). Hypothetical and real incentives in the ultimatum game and Andreoni's public goods game: An experimental study. *Eastern Economic Journal, 33*(4), 491–510.

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives, 25*(4), 191–210. http://dx.doi.org/10.1257/jep.25.4.191.

Heck, D. W., & Moshagen, M. (2016). RRreg: Correlation and regression analyses for Randomized Response data. R package version 0.6.1. Retrieved from https://cran.r-project.org/package=RRreg.

Heck, D. W., Wagenmakers, E.-J., & Morey, R. D. (2015). Testing order constraints: Qualitative differences between Bayes factors and normalized maximum likelihood. *Statistics & Probability Letters, 105*, 157–162. http://dx.doi.org/10.1016/j.spl.2015.06.014.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2-3), 61–83. http://dx.doi.org/10.1017/S0140525X0999152X.

Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences, 24*(3), 383–451.

Hilbig, B. E., Moshagen, M., & Zettler, I. (2015). Truth will out: Linking personality, morality, and honesty through indirect questioning. *Social Psychological and Personality Science, 6*(2), 140–147. http://dx.doi.org/10.1177/1948550614553640.

Hilbig, B. E., Thielmann, I., Hepp, J., Klein, S. A., & Zettler, I. (2015). From personality to altruistic behavior (and back): Evidence from a double-blind dictator game. *Journal of Research in Personality, 55*, 46–50. http://dx.doi.org/10.1016/j.jrp.2014.12.004.

Hilbig, B. E., & Zettler, I. (2015). When the cat's away, some mice will play: A basic trait account of dishonest behavior. *Journal of Research in Personality, 57*, 72–88. http://dx.doi.org/10.1016/j.jrp.2015.04.003.

Hoffman, E., McCabe, K. A., Shachat, K., & Smith, V. (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior, 7*(3), 346–380. http://dx.doi.org/10.1006/game.1994.1056.

Hoffmann, A., & Musch, J. (2015). Assessing the validity of two indirect questioning techniques: A Stochastic Lie Detector versus the Crosswise Model. *Behavior Research Methods*, 1–15. http://dx.doi.org/10.3758/s13428-015-0628-6.

Holm, H. J., & Nystedt, P. (2008). Trust in surveys and games–A methodological contribution on the influence of money and location. *Journal of Economic Psychology, 29*(4), 522–542. http://dx.doi.org/10.1016/j.joep.2007.07.010.

Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology, 32*(5), 865–889. http://dx.doi.org/10.1016/j.joep.2011.05.007.

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. New York, NY, US: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139087759.

Lensvelt-Mulders, G. J. L. M., Hox, J. J., Van Der Heijden, P. G. M., & Maas, C. J. M. (2005). Meta-analysis of Randomized Response research: Thirty-five years of validation. *Sociological Methods & Research, 33*(3), 319–348. http://dx.doi.org/10.1177/0049124104268664.

List, J. A. (2007). On the interpretation of giving in dictator games. *Journal of Political Economy, 115*(3), 482–493.

Lönnqvist, J.-E., Verkasalo, M., & Walkowitz, G. (2011). It pays to pay—Big Five personality influences on cooperative behaviour in an incentivized and hypothetical prisoner's dilemma game. *Personality and Individual Differences, 50*(2), 300–304. http://dx.doi.org/10.1016/j.paid.2010.10.009.

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E.-J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review, 23*(1), 103–123. http://dx.doi.org/10.3758/s13423-015-0947-8.

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods, 42*(1), 42–54. http://dx.doi.org/10.3758/BRM.42.1.42.

Moshagen, M., Hilbig, B. E., Erdfelder, E., & Moritz, A. (2014). An experimental validation method for questioning techniques that assess sensitive issues. *Experimental Psychology, 61*(1), 48–54. http://dx.doi.org/10.1027/1618-3169/a000226.

Moshagen, M., Hilbig, B. E., & Musch, J. (2011). Defec-

tion in the dark? A randomized response investigation of cooperativeness in social dilemma games. *European Journal of Social Psychology, 41*(5), 638–644. http://dx.doi.org/10.1002/ejsp.793.

Moshagen, M., & Musch, J. (2012). Surveying multiple sensitive attributes using an extension of the randomized-response technique. *International Journal of Public Opinion Research, 24*(4), 508–523. http://dx.doi.org/10.1093/ijpor/edr034.

Moshagen, M., Musch, J., & Erdfelder, E. (2012). A stochastic lie detector. *Behavior Research Methods, 44*(1), 222–231. http://dx.doi.org/10.3758/s13428-011-0144-2.

Naef, M., & Schupp, J. (2009). *Measuring trust: Experiments and surveys in contrast and combination*. SOEP-paper No. 167. http://dx.doi.org/10.2139/ssrn.1367375.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing* (Vol. 124, p. 125). Vienna, Austria.

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature, 489*(7416), 427–430. http://dx.doi.org/10.1038/nature11467.

Reips, U.-D. (2002a). Internet-based psychological experimenting: Five dos and five don'ts. *Social Science Computer Review, 20*(3), 241–249. http://dx.doi.org/10.1177/08939302020003002.

Reips, U.-D. (2002b). Standards for Internet-based experimenting. *Experimental Psychology, 49*(4), 243–256. http://dx.doi.org/10.1026//1618-3169.49.4.243.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*(2), 225–237. http://dx.doi.org/10.3758/PBR.16.2.225.

Statistisches Bundesamt. (2012). *Geburten in Deutschland*. Retrieved from https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/Bevoelkerungsbewegung/BroschuereGeburtenDeutschland0120007129004.pdf?__blob=publicationFile.

Thielmann, I., & Hilbig, B. E. (2015). Trust: An integrative review from a person-situation perspective. *Review of General Psychology, 19*(3), 249–277. http://dx.doi.org/10.1037/gpr0000046.

Thielmann, I., Hilbig, B. E., Zettler, I., & Moshagen, M. (in press). On measuring the sixth basic personality dimension: A comparison between HEXACO Honesty-Humility and Big Six Honesty-Propriety. *Assessment*. http://dx.doi.org/10.1177/1073191116638411.

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology, 60*(3), 158–189. http://dx.doi.org/10.1016/j.cogpsych.2009.12.001.

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association, 60*(309), 63–69. http://dx.doi.org/10.2307/2283137.

Yu, J.-W., Tian, G.-L., & Tang, M.-L. (2007). Two new models for survey sampling with sensitive characteristic: design and analysis. *Metrika, 67*(3), 251–263. http://dx.doi.org/10.1007/s00184-007-0131-x.

Zhang, L., & Ortmann, A. (2014). The effects of the take-option in dictator-game experiments: A comment on Engel's (2011) meta-study. *Experimental Economics, 17*(3), 414–420. http://dx.doi.org/10.1007/s10683-013-9375-7.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics, 13*(1), 75–98. http://dx.doi.org/10.1007/s10683-009-9230-z.