Open access • Journal Article • DOI:10.1257/AER.103.1.532

Preferences for Truthfulness: Heterogeneity among and within Individuals

Rajna Gibson, Carmen Tanner, Alexander F. Wagner

Institutions: Swiss Finance Institute, Harvard University

Published on: 01 Feb 2013 - The American Economic Review (American Economic Association)

Article

Preferences for Truthfulness: Heterogeneity among and within Individuals

GIBSON BRANDON, Rajna Nicole, TANNER, Carmen, WAGNER, Alexander F

Abstract

We conduct an experiment assessing the extent to which people trade off the economic costs of truthfulness against the intrinsic costs of lying. The results allow us to reject a type-based model. People's preferences for truthfulness do not identify them as only either "economic types" (who care only about consequences) or "ethical types" (who care only about process). Instead, we find that preferences for truthfulness are heterogeneous among individuals. Moreover, when examining possible sources of intrinsic costs of lying and their interplay with economic costs of truthfulness, we find that preferences for truthfulness are also heterogeneous within individuals.

Reference

GIBSON BRANDON, Rajna Nicole, TANNER, Carmen, WAGNER, Alexander F. Preferences for Truthfulness: Heterogeneity among and within Individuals. *American Economic Review*, 2013, vol. 103, no. 1, p. 532-548

DOI : 10.1257/aer.103.1.532

American Economic Review 2013, 103(1): 532–548
http://dx.doi.org/10.1257/aer.103.1.532

Preferences for Truthfulness: Heterogeneity among and within Individuals[†]

By Rajna Gibson, Carmen Tanner, and Alexander F. Wagner*

Standard economic models of self-interested utility maximization, which emphasize the role of consequences in determining agents' actions, predict a grim inexorability to all economic systems. These models are based on an assessment of humans as self-interested agents who behave dishonestly for cogent reasons. These hypothetical persons prioritize the outcomes of their actions and forgo materially beneficial lying only if strategic or reputational considerations arise. Some researchers, such as Bhide and Stevenson (1990), assert that these reputational forces are often weak, implying that honesty simply does not seem to pay.

Examples of disastrous dishonesty based on such self-interest abound in the corporate world. Deliberate deception has augmented the economic effects of regulatory failure, of a deteriorating macro-economy, and of inadequate models, in, for example, the subprime crisis.

Yet, truthfulness also appears to prosper in society. Whistleblowers often jeopardize their careers and friendships when they truthfully reveal the wrongdoing of their companies. Some CEOs are regarded as particularly virtuous (Treviño and Brown 2004). Numerous journalists risk their lives to report the truth about political repression, economic crimes, and human rights violations.

To explain otherwise puzzling behavior both in the field and in experiments, several authors have proposed the idea that some people experience intrinsic costs when they lie. For example, in a cheap-talk sender-receiver game, Gneezy (2005) found that many subjects told the truth.[1] Of various possible explanations for this result, he

* Gibson: Swiss Finance Institute–University of Geneva, 40 Bd Pont d'Arve, CH-1211 Genève 4, Switzerland, and Geneva Finance Research Institute (GFRI) (e-mail: rajna.gibson@unige.ch); Tanner: Department of Banking and Finance, University of Zurich, Plattenstrasse 32, CH-8032 Zurich, Switzerland (e-mail: carmen.tanner@bf.uzh.ch); Wagner: Department of Banking and Finance, University of Zurich, Plattenstrasse 14, CH-8032, Zurich, Switzerland, Swiss Finance Institute–University of Zurich, Harvard University, and CEPR (e-mail: alexander.wagner@bf.uzh.ch).
We thank three anonymous referees for excellent remarks that have greatly improved the paper. Rainer Winkelmann, who went far beyond the call of duty in helping us clarify the empirical framework and its relation to our preferred preference specification, deserves our special thanks. We also thank Peter Bossaerts, Chiddi Chidambaran, Alain Cohn, François Degeorge, Ernst Fehr, Gerlinde Fellner, Laurent Frésard, Michelle Sovinsky, Susanne Neckermann, Pawel Polak, Bernard Raffournier, Jean-Charles Rochet, Meir Statman, and Mei Wang for fruitful discussions and comments. Useful comments were provided by seminar participants at the European Finance Association meetings, the International Conference on Psychology, the Conference of the Society of Judgment and Decision Making, the European Association of Experimental Social Psychology, la Universidad Carlos III, Fudan University, the University of Fribourg, Northwestern University, Rutgers University, the University of St. Gallen, and the University of Zurich. Part of this work was conducted while Gibson was visiting UCLA. We thank the Research Priority Program Finance and Financial Markets of the University of Zurich, the NCCR FINRISK, the Swiss Finance Institute, and the Swiss National Science Foundation (Grant #PP001-102845) for financial support. Nicolas Berkowitsch provided excellent research assistance. We thank Marcel Morf, who programmed the first version of the code for the experiments.

[†] To view additional materials, visit the article page at http://dx.doi.org/10.1257/aer.103.1.532.

[1] Similar results on truth telling have been obtained in other studies (Evans, et al. 2001; Sánchez-Pagés; and Vorsatz 2007). Only a few researchers, such as Baiman and Lewis (1989), have found that people will lie for even just a tiny monetary payoff. See the edited volume by Zak (2008) for numerous additional examples.

inferred that the most plausible was that "people have non-consequential preferences in which they treat the same monetary outcome differently, depending on the process that leads up to it" (p. 392). Moreover, in Gneezy's interpretation, "different people weigh these preferences differently" (p. 392). That is, a model in which agents exhibit (continuously) heterogeneous preferences for truthfulness could explain his data.

Because Gneezy's experiment was set in a strategic context, social preferences may also have been active. Therefore, Gneezy also emphasized the joint relevance of process-dependent preferences and of consequences to oneself and to others.[2] But it is precisely because both lying aversion and social preferences operated in his experiment that the two channels were difficult to isolate. Indeed, Hurkens and Kartik (2009, p. 180) showed that Gneezy's (2005) empirical observations were consistent with the "hypothesis that people are one of two kinds: either a person will never lie, or a person will lie whenever she prefers the outcome obtained by lying over the outcome obtained by telling the truth." Based on this existing evidence, it is, therefore, possible that the world is populated, in the spirit of type-based models such as that of Koford and Penno (1992), by exactly two fixed types: economic types and ethical types (in Gneezy's terminology). Alternatively, these two types can, respectively, be characterized as consequentialists (who care about consequences to themselves and to others, but not about the process by which these consequences are achieved) and as nonconsequentialists (who care only about the process, but not about consequences).

The type–based model and the model with heterogeneous preferences for truthfulness lead to very different implications, particularly for agent selection and incentive design. Therefore, it is important to determine which of these two models offers a more accurate description of reality.

To address this question, we conducted a decision-theoretic laboratory experiment in which each participant was placed in the situation of a CEO who had to announce his or her firm's earnings to a passive market. The participants were informed of the true level of earnings. They were also told that falsely reporting higher earnings was legal and would lead to higher actual payoffs than honestly announcing the lower earnings. We considered that economic types would always lie in our experiment because truthfulness was designed to be economically costly; we considered that ethical types would always tell the truth. If, by contrast, individuals varied continuously in the extent to which they were driven by preferences for truthfulness, they would trade off the economic costs of truthfulness with the costs of lying; those with intermediately strong preferences for truthfulness would exhibit the most changes in behavior as economic costs changed.

The simplicity of our experimental setup—involving a decision-making situation with no counterparties—allowed us to isolate motivations for truthtelling that are nonstrategic and not driven by social preferences; it permitted us to sidestep issues that occur in strategic contexts. Moreover, in our experiment, we observed individuals' behaviors. This setup enabled us to provide evidence, stronger than that developed in existing works, regarding heterogeneity in preferences for truthfulness.

[2] A long-standing literature considers the role of preferences that depend on process and/or on consequences for others; it recognizes that people do not necessarily maximize utility according to the material consequences of their actions. For example, Rabin (1995) demonstrated how fairness considerations can explain why people are willing to reward or punish others even when this requires a sacrifice of their own well-being.

We observed that, in a situation where the standard economic model predicts that everybody will lie, 32 percent of the participants chose not to do so, thus forgoing a larger variable compensation. Importantly, the aggregate percentage of truthtellers decreased as the costs of truthfulness increased. Our individual-level regressions imply that the marginal effect of a cost increase on the probability of an individual's telling the truth is significantly negative, even after controlling for various demographic and psychological factors. These results are at odds with the type-based model but are consistent with a model that posits heterogeneous preferences for truthfulness.

Our primary contribution, therefore, is to provide evidence for the notion that people occupy a spectrum of preferences for truthfulness rather than only two opposite positions.[3] As a secondary contribution, we examine potential sources of the heterogeneity in preferences for truthfulness. Tendencies towards impression management and self-deception offer no explanatory power; however, one measure of a source of intrinsic costs of lying, an index of "protected values of truthfulness," seems to organize the data well. We also find substantial evidence of nonseparability between this measure of intrinsic costs of lying and economic costs of truthfulness in the utility function. In other words, total preferences for truthfulness display heterogeneity not only among, but also within, individuals. We do not have adequate measures of other possible sources of intrinsic costs of lying, including, in particular, measures of expressive preferences. Therefore, we acknowledge that other preference formulations could potentially explain our empirical evidence.

Section I presents the basic trade-off and the hypotheses. Section II describes the experiment. Section III discusses the main results. Section IV explores possible sources of intrinsic costs of lying and their interaction with economic costs of truthfulness. Section V concludes.

I. The Trade-off

Consider an agent who decides whether to tell the truth, $T = 1$, or to lie, $T = 0$. Lying, the agent receives a certain income m. There are economic costs of stating the truth, for which we use the term *ECOST*. The agent receives funds *m-ECOST* when he tells the truth. We model *preferences for truthfulness* by positing that the agent also experiences *total costs of lying*, C_i. (For the moment, C_i is given. We discuss in Section IV how these total costs of lying may arise from the interplay between the intrinsic costs of lying and the extrinsic economic costs of truthfulness.) If types are continuous, C_i can take on any value, positive or negative. By contrast, in the two-type model, there are only "ethical types" who have $C_i = \infty$ and "economic types" who have $C_i = 0$. Let the global utility function be defined as

$$V_i(T) = \begin{cases} b(m - ECOST) & \text{if } T = 1 \\ bm - C_i & \text{if } T = 0, \end{cases} \tag{1}$$

[3] Gneezy, Imas, and Madarász (2012) found that participants who lied in a sender-receiver game were more likely to later donate to charity than those who chose to tell the truth. This also suggests that there are not simply only ethical types and economic types.

where T is the choice variable.[4] For simplicity, and because wealth effects are unlikely in our experiment, we assume the agent has a constant marginal utility of money $b > 0$. We also assume that all participants have the same b.[5]

The difference between the utilities of truthtelling and of lying is given by

$$Y_i^* = C_i - bECOST. \tag{2}$$

An individual exhibits truthfulness when $Y_i^* > 0$. This implies that truthfulness can, in this framework, arise as optimal behavior only if there is a positive total cost of lying. While social preferences are known to contribute to behavior (e.g., Fehr and Fischbacher 2002, 2003), our experiment is designed to eliminate any role for altruism, reciprocity, guilt aversion (Charness and Dufwenberg 2006), and related factors, as well as any role for strategic concerns that might arise with repeated interaction.

Consider now a population of individual decision makers (whose distribution of C_i is not known), each of them facing various economic costs of truthfulness. A type-based model, such as that of Koford and Penno (1992), implies that ethical types, with their overwhelming preferences for truthfulness, would always choose $T = 1$, and this choice would be invariant to *ECOST*. Conversely, economic types would always lie when profitable. (At $ECOST = 0$, they would perceive no advantage or disadvantage to either telling the truth or lying; but at all other levels of economic costs of truthtelling, the utility difference Y_i^* would be negative.) Aggregating across the population of individuals, this implies the following hypothesis:

HYPOTHESIS TYP (Type-based model): *The fraction of the population telling the truth remains constant across varying economic costs of truthfulness.*

By contrast, in the model based on heterogeneous preferences for truthfulness, where C_i varies continuously throughout the population, varying economic costs would lead some individuals with intermediate total lying costs to change their behavior. Higher economic costs of truthfulness would then make it less likely that an individual would tell the truth. Thus, we have the alternative hypothesis reflecting Gneezy's conjecture:

HYPOTHESIS HET (Model based on heterogeneous preferences for truthfulness): *The fraction of the population telling the truth varies with economic costs of truthfulness.*

In Section III, we test these two hypotheses using aggregate behavioral data. We also specify an empirical model for individual choice to test the corresponding underlying predictions regarding the marginal effect of economic costs of truthfulness on individual choice.

[4] Truthfulness here is a matter of preference. Alternatively, we could posit a constraint involving a need to maintain a minimum level of truthtelling. Within this simple context, the two formulations are identical. Rabin (1995) showed that moral preferences and moral constraints can result in different incentives for information collection.

[5] It is standard to assume that, abstracting from the preference feature of interest (for example, inequity aversion), all participants have equal marginal utility of money. See, for example, Fehr and Schmidt (1999).

II. Experimental Method[6]

We are interested in situations requiring a choice between telling the truth and telling a lie, in which the former decision involves an economic sacrifice. As our context, we chose accounting earnings management (henceforth called "earnings management").[7] This situation illuminates a real-life conflict: management's variable compensation is frequently tied to stock price performance, which in turn often hinges on earnings announcements. We envisioned a framework in which earnings management would be understood to be legal (for example, within GAAP rules) although explicitly self-interested and dishonest—a decision-making problem focused exclusively on the managerial choice. We required the recipient (the market, played by the computer) to accept passively all financial statements. The advantage of this approach is that, due to the absence of strategic interactions, we have been able to isolate, at least better than in the real world, factors influencing individuals' choices, without monitoring the participants' thoughts regarding the behavior of other players.[8]

A. *Participants and Procedure*

A total of 261 participants (median age: 23 years) took part in this online experiment. We recruited participants from undergraduate classes at the University of Zurich. Fifty percent of the participants were economics and finance students, 40 percent psychology students, and 10 percent students from other fields. Forty-two percent were women, and 58 percent were men (distributed across the fields). All participants were told at the outset that anonymity was ensured.[9] They were first asked to respond to a few demographic questions and to read some basic instructions. They were informed that they would individually receive a payment, CHF 8, for their completed participation in the study, and an additional payment that depended on their decisions. After having demonstrated their understanding of the (unlabeled) tasks and of the rules of the experiment, the participants completed, in randomized orders, the three main parts of the experiment: (i) the truthtelling task, (ii) the effort task, and (iii) the measurement of various controls and potential

[6] The full set of instructions is available in the online Appendix. The experiment included other tasks whose content and results are not described here for space reasons.

[7] Accounting earnings management occurs "when managers use judgment in financial reporting and in structuring transactions to alter financial reports to either mislead some stakeholders about the underlying economic performance of the company or to influence contractual outcomes that depend on reported accounting numbers" (Healy and Wahlen 1999). Accounting earnings management can be viewed as a form of lying, which is defined as "a statement that one knows to be false" (Grover 2005). Nelson, Elliott, and Tarpley (2003) provide many examples collected from auditors.

[8] In particular, despite the intuitive appeal and real-world relevance of the strategic games employed in Gneezy (2005) and in Hurkens and Kartik (2009), these games come with some interpretational challenges. For instance, in sender-receiver games, even telling the truth can be deceptive, because the sender may hope that the receiver will not believe the true message that is sent (Sutter 2009). Additionally, Rode (2010) found that decision makers were significantly less trusting in a competitive context than in a cooperative context.

[9] At the time of the experiment, the exchange rate was about US $1 = CHF 1.15. Most participants received payment one week after the experiment. For this purpose, each participant received, before the experiment, a code, based on which the experimenter prepared an envelope containing the earnings. Participants received the sealed envelopes by indicating their personal codes. It is, therefore, unlikely that a desire to appear honest affected the participants' behavior systematically. Ariely, Bracha, and Meier (2009) document how publicly displayed monetary incentives can be less effective in promoting prosocial behavior than privately displayed incentives.

proxies for intrinsic costs of lying. Finally, all the participants were paid. For simplicity, we describe the procedure for one of the randomized orders of tasks.

(i) In the truthtelling task, each participant was placed in the situation of a CEO who had to announce earnings per share for the previous quarter. The participants were told that the variable component of their salaries would depend on the earnings they announced. They were also told that the market currently anticipated the announcement of 35 cents per share as earnings, but that the true earnings were 31 cents per share. The participants were told that they could announce earnings of 35 cents per share while remaining within legal accounting limits,[10] and that the decision would be solely theirs. They were also informed that they would be paid an amount based on the CEO compensation (according to their decisions). This additional experimental payoff would be converted into real money at the rate of CHF 100,000 = CHF 0.5. Importantly, participants earned less when choosing to tell the truth.

The participants were then told they would have to announce their financial statements that day. The truthtelling task questionnaire follows:

Which earnings will you announce?

__ 31 cents per share – In this case, your compensation will be CHF 60,000 (CHF 0.30).
__ 35 cents per share – In this case, your compensation will be CHF 300,000 (CHF 1.50).

__ 31 cents per share – In this case, your compensation will be CHF 120,000 (CHF 0.60).
__ 35 cents per share – In this case, your compensation will be CHF 300,000 (CHF 1.50).

__ 31 cents per share – In this case, your compensation will be CHF 180,000 (CHF 0.90).
__ 35 cents per share – In this case, your compensation will be CHF 300,000 (CHF 1.50).

__ 31 cents per share – In this case, your compensation will be CHF 240,000 (CHF 1.20).
__ 35 cents per share – In this case, your compensation will be CHF 300,000 (CHF 1.50).

__ 31 cents per share – In this case, your compensation will be CHF 300,000 (CHF 1.50).
__ 35 cents per share – In this case, your compensation will be CHF 300,000 (CHF 1.50).

A few questions served as a manipulation check to verify that participants distinguished between the 31 and 35 cent options. The participants were asked, using a five-point scale ranging from -2 to $+2$, the extent to which they judged announcing 31 cents as *dishonest* versus *honest*, *manipulative* versus *not manipulative*, *short term–oriented* versus *long term–oriented*, and *associated with personal benefits* versus *associated with personal costs*. The same was also done for the 35-cent announcement option.

(ii) Participants engaged in a simple calculation (effort) task.

[10] Therefore, risk preferences of individuals did not matter, as their choices were not based on the trade-off between the expected benefits and costs of committing a crime.

(iii) We then measured, as potential sources of intrinsic costs of lying (a term we introduce formally in Section IV), their tendencies towards impression management and self-deception, and their levels of protected values. Moreover, we also measured their altruistic concerns.

After the experiment, the participants anonymously received their payments of CHF 8 plus their earnings. The average total payment was slightly less than CHF 30.5.[11]

B. *Variables of Interest*

TRUTHFUL CHOICE.—This represented the dependent variable in the truthtelling task, coded as a binary variable that took on the value of 1 if a participant chose to announce earnings of 31 cents (the honest option), and the value of 0 if a participant announced 35 cents (the dishonest option).

ECOST.—This was a within-participants variation. Economic costs of truthfulness derived from the amount of money a participant forfeited by announcing 31 cents. The *ECOST* variable took on values from CHF 0 to CHF 1.20 ($= 1.50 - 0.30$), in increments of 30 cents.

IMPRESSION MANAGEMENT and SELF DECEPTION.—Using the standard Deception Scales (PDS) of Paulhus (1984) in the German version of Musch, Brockhaus, and Bröder (2002), we measured individuals' tendencies to give socially desirable responses. These tendencies come in two distinct forms: a tendency to deceive others (impression management) and a tendency to deceive oneself (self-deception). Both are expected to be positively related to intrinsic costs of lying. Accordingly, we coded two variables, *EXTDECEIT* and *SELFDECEIT*. We scaled the measures to be between 0 and 1. Participants who exhibited more socially acceptable responses scored higher on both scales.

PROTECTED VALUES (*PV*).—The extent to which participants held truthfulness as a protected value and, therefore, felt committed to truthtelling was another source of intrinsic costs of lying. To measure this source, we used an index developed by Tanner, Ryf, and Hanselmann (2009), the details of which are available in the online Appendix. This index took on a value between 0 (for an individual with no protected values) and 6 (for an individual with maximum protected values). The internal consistency of this scale, as assessed by Cronbach's α, was very satisfactory ($\alpha = 0.86$).[12]

ALTRUISTIC CONCERNS.—We asked participants the extent to which they believed that announcing 35 cents had consequences for other stakeholders ($-2 =$ *hurting other stakeholders* to $+2 =$ *not hurting other stakeholders*). Of course, within the strict confines of the experiment, there were no such consequences. Nonetheless, this variable was a relevant control for any altruistic preferences or

[11] This amount includes payment for other tasks in the full experiment.

[12] Cronbach's Alpha is a measure of the reliability and the internal consistency of an instrument. The measure ranges from 0 to 1 and will generally increase when the correlations between the items increase.

TABLE 1—MANIPULATION CHECKS

	31 cents	35 cents	*t*-test
Honest vs. dishonest	1.61	−1.17	26.47
Nonmanipulative vs. manipulative	1.39	−1.14	22.70
Personal financial loss vs. personal financial gain	0.93	−1.25	19.50
Long-term view vs. short-term view	0.99	−1.1	18.53

Notes: Participants answered questions that asked for their assessments of announcing 31 and 35 cents, respectively. These questions were on a −2 to +2 scale. After reordering (the direction of the scale varies between questions), a value of +2 indicates that the action was seen as honest, nonmanipulative, associated with a personal financial loss, and based on a long-term view.

fairness concerns of the participants which might confound our inferences. Answers to this question were coded as the variable 35*HURTS*.

DEMOGRAPHIC CONTROL VARIABLES.—SEX was equal to 1 for female participants and to 0 for male participants. *AGE* was equal to each participant's age in completed years. *PSYCHOLOGY* was equal to 1 for psychology students ("psychologists") and to 0 otherwise. *OTHER* was equal to 1 for participants from fields other than psychology and economics and to 0 otherwise. *ECONOMICS* was the omitted category.

III. Main Results

A. *Descriptive Evidence*

We first confirm, through a manipulation check, that the participants generally understood the announcement of 31 cents to be the honest, nonmanipulative action that would lead to a personal loss, while the opposite was true of the announcement of 35 cents. (See Table 1.)

Table 2 allows a first look at the choices the participants made in the experiment. In approximately 42 percent of cases (32 percent when omitting the free-truth situation), participants chose to announce low earnings, that is, chose not to engage in earnings management. By telling the truth, those participants opted to suffer, on average, effective monetary losses of 11 percent of the maximum total amount they could have earned in the truthtelling task or 27 percent of the variable amount they could have earned above the guaranteed payout.

The fact that a large proportion of the participants reported the truth, even when the conditions opposed it, is consistent with the notion that many individuals have positive total costs of lying. By contrast, this finding is inconsistent with the standard economic model.

B. *Reactions to Economic Costs of Truthfulness*

In this section, our primary goal is to test the implications of the type-based model, which posits that the (participant) population consisted only of "economic types" and "ethical types," against the implications of the model based on heterogeneous preferences for truthfulness.

TABLE 2—BEHAVIOR ACROSS ECONOMIC COSTS OF TRUTHFULNESS

ECOST = CHF 0	*ECOST* = CHF 0.3	*ECOST* = CHF 0.6	*ECOST* = CHF 0.9	*ECOST* = CHF 1.2	Total	Total except *ECOST* = 0
82.0	52.1	31.4	23.0	21.1	41.9	31.9

Note: This table presents the percentages of participants announcing 31 cents of earnings per share across the various economic cost of stating the truth (*ECOST*) conditions.

Table 2, showing aggregate data, reveals substantial variation in the participants' responses as the economic costs of truthfulness changed: with higher economic costs, the percentage of participants telling the truth was lower. This is inconsistent with the type-based model's prediction in Hypothesis TYP, according to which the fraction of participants who told the truth would have remained constant. Formally, a χ^2-test strongly rejects the hypothesis that there is a fixed fraction of "ethical" types who always tell the truth and a fixed fraction of "economic types" who always lie, with nobody differing from these two types. (This is true not only of the specific version postulated by Hurkens and Kartik (2009), where exactly half of the population always tell the truth and the other half always lie, but also for any other fraction between 0 and 1.) When there was no economic cost of truthfulness, 18 percent of the participants still chose the earnings-management solution. This can be explained by recognizing that the model based on heterogeneous preferences for truthfulness allows agents to have a negative total cost of lying. (The manipulation check confirms that this group of people perceived 35 cents as the less honest option.)

To investigate statistically the influence of the economic costs of truthfulness on individual behavior, we estimate a discrete choice/random utility model (e.g., King 1998; Wooldridge 2006). From equation (2), adding a stochastic error and rearranging terms, each participant i's $(i = 1, \ldots, 261)$ latent utility difference between truthtelling and lying at direct economic $ECOST_j$ is given by

$$Y_{ij}^{*} = C_i - bECOST_j + \varepsilon_{ij}. \tag{3}$$

Under utility maximization, an observed realization of *TRUTHFUL CHOICE*, T_{ij}, is related to Y_{ij}^{*} by the following mechanism:

$$T_{ij} = \begin{cases} 1 & \text{if } Y_{ij}^{*} \geq 0 \\ 0 & \text{if } Y_{ij}^{*} < 0. \end{cases} \tag{4}$$

In line with standard practice, we assume that ε is independent of the explanatory variables $\mathbf{X}$. By assuming that ε has the logistic distribution, one obtains the logit model, which is the main specification on which we focus. After relabeling and combining coefficients,

$$\Pr(T_{ij} = 1 \,|\, \mathbf{X}) = \Lambda[\beta_0 + \beta_E ECOST_j], \tag{5}$$

where $\Lambda(\cdot)$ is the logistic cumulative distribution function.[13] The coefficient vector is estimated by maximum likelihood. If $\hat{\beta}_E < 0$, participants react negatively overall to economic costs. The coefficient $\hat{\beta}_0$ here gives the average C_i in the data for zero economic costs of truthfulness. (In Section IV, we explore sources of variation in C_i and we discuss that, besides the direct effect of reducing the attractiveness of truthfulness, *ECOST* may have an indirect effect through the total costs of lying.) The standard errors correct for possible serial correlation and heteroskedasticity by clustering at the individual level. (Recall that participants went through all five economic cost situations.)

Column 1 of Table 3 shows the results of this analysis. Consistent with the fact that many individuals did, in fact, tell the truth, the constant term is positive. Importantly, *ECOST* is a highly significant determinant of the relative attractiveness of truthfulness and lying for an individual. Indeed, the implied marginal effect of *ECOST* is powerful: A 30-cent increase in *ECOST* was associated with a 16.9 percent decrease in truthtelling.[14] Together with the observations made in Table 2, this finding supports Hypothesis HET.

Column 2 of Table 3 adds individual-level controls. The main result for *ECOST* remains unchanged. We observe some interesting additional findings. First, women appeared to be more likely to tell the truth, as did students in fields outside of psychology and economics. Second, given the decision-theoretic nature of the experiment, altruistic and distributional concerns, as well as attempts to live up to others' expectations so as to avoid guilt, should not have affected behavior in this experiment. Yet, we note that 35*HURTS* does enter significantly in this baseline regression. This suggests one of two possibilities. Either participants' altruistic concerns drove behavior, or 35*HURTS* was correlated with some general differences in preferences that were, in turn, correlated with intrinsic costs of lying and, thus, reflected variation in a variable omitted in column 2. As Section IV shows, the latter explanation is corroborated by the data.

The conclusions we draw from these main results are simple but important. Hurkens and Kartik (2009) demonstrated that Gneezy's (2005) data would be consistent with a population of pure opportunists, who always lie, and pure ethical types, who always tell the truth (as in Koford and Penno 1992). However, the present evidence of changeability in truthtelling behavior and of significant sensitivity to economic costs associated with truthfulness rejects Hypothesis TYP and is in direct contrast to the implications of a type-based model.

IV. Sources of Heterogeneity in Total Costs of Lying

We have established that the participants in our experiment showed more variation in their total costs of lying than if they had belonged to one of just two fixed extreme types. In this section, we expand on these findings by considering various potential sources of the variation among individuals in total costs of lying. Moreover, this

[13] If ε is normally distributed, one obtains the probit model. As is typical in econometric applications, the two models yield virtually identical inferences.

[14] In the real world, managers are indeed faced with substantial cross-sectional and time-series variation in the economic cost of truthtelling. Our results are consistent with findings by Bergstresser and Philippon (2006), who showed that the use of discretionary accruals to manipulate reported earnings was more pronounced at firms where CEO compensation depended more on the stock price.

Table 3—Determinants of Earnings Management Behavior

	(1)	(2)	(3)	(4)	(5)
Economic cost of no earnings management (*ECOST*)	−2.35*** (0.17)	−2.50*** (0.18)	−2.75*** (0.20)	−6.10*** (1.38)	−5.22*** (1.43)
Sex (1: Female, 0: Male)	—	0.46* (0.26)	0.30 (0.28)	0.30 (0.28)	0.39 (0.30)
Age (years)	—	−0.02 (0.03)	−0.03 (0.03)	−0.03 (0.03)	−0.01 (0.03)
Psychology (1: Yes, 0: No)	—	0.25 (0.29)	0.03 (0.31)	0.05 (0.31)	−0.13 (0.34)
Other studies (1: Yes, 0: No)	—	0.66* (0.38)	0.24 (0.40)	0.27 (0.40)	0.07 (0.44)
Altruistic concerns (35*HURTS*)	—	0.41*** (0.11)	0.22** (0.11)	0.12 (0.11)	−0.04 (0.16)
Impression management tendency (*EXTDECEIT*)	—	—	0.50 (0.97)	−0.07 (0.87)	0.66 (1.13)
Self-deception tendency (*SELFDECEIT*)	—	—	0.02 (1.03)	0.08 (0.97)	−1.07 (1.25)
Protected values (*PV*)	—	—	0.73*** (0.14)	0.36*** (0.13)	0.45** (0.19)
35*HURTS* × *ECOST*	—	—	—	0.18 (0.21)	0.35 (0.24)
EXTDECEIT × *ECOST*	—	—	—	1.18 (1.59)	0.14 (1.61)
SELFDECEIT × *ECOST*	—	—	—	−0.66 (1.74)	0.79 (1.78)
PV × *ECOST*	—	—	—	0.75*** (0.22)	0.57** (0.25)
Constant	1.00*** (0.11)	0.88 (0.66)	−1.56 (1.16)	0.16 (1.11)	−0.61 (1.26)
Observations	1,305	1,305	1,305	1,305	1,044
Number of participants	261	261	261	261	261
Pseudo R^2	0.15	0.19	0.25	0.26	0.19
Pseudo log-likelihood	−757.6	−719.7	−667.4	−655.7	−531.8
Likelihood-ratio test statistic (χ^2, *p*-value)	259.6 (<0.01)	335.4 (<0.01)	440.0 (<0.01)	463.4 (<0.01)	243.6 (<0.01)
Wald test statistic (χ^2, *p*-value)	183.9 (<0.01)	211.0 (<0.01)	222.8 (<0.01)	204.8 (<0.01)	107.1 (<0.01)

Notes: This table presents coefficients of logit regressions. The dependent variable is *TRUTHFUL CHOICE*. The explanatory variables are described in the text. Columns 1 to 4 use data from all *ECOST* situations. Column 5 omits the free truth (*ECOST* = CHF 0) situation. Robust standard errors, obtained by clustering at the individual level, appear in parentheses below coefficient estimates.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

extension allows us to provide further evidence refuting the type-based model; in particular, we document that there is also heterogeneity *within* individuals (across situations) in total costs of lying.

A. *Enhanced Model*

We consider two constituent sources of heterogeneity in total costs of lying. First, these costs are driven by individuals' intrinsic costs of lying ($ICOL_i$). Second, we

also allow the situation, that is, the economic costs of truthfulness or the extrinsic incentives for lying, $ECOST_j$, to influence the total costs of lying. Adjusting notation, we posit that total costs of lying may vary both among and within individuals, and we now write $C_{ij} = C_{ij}(ICOL_i, ECOST_j)$. Since all participants encountered the same *ECOST* situations, the evidence on heterogeneous total costs of lying provided in Section III necessarily implies that there is heterogeneity in $ICOL_i$. Next, the intrinsic costs of lying and the economic incentives for lying (economic costs of truthfulness) may enter C_{ij} separably or nonseparably. Indeed, whether intrinsic preferences and extrinsic incentives interact in determining total preferences for a certain action has implications that extend beyond the scope of the current study; see, for example, Bowles and Polanía-Reyes (2012) for a discussion of how incentives to contribute to public goods may affect social preferences. In order to capture both possibilities, we consider, for parsimony, a simple parametric specification:

$$(6)\quad C_{ij}(ICOL_i, ECOST_j) = \gamma_0 + \gamma_1 ICOL_i + \gamma_2 ECOST_j + \gamma_3 ICOL_i ECOST_j,$$

implying

$$(7)\quad Y^*_{ij} = C_i - bECOST_j + \varepsilon_{ij}$$

$$= \gamma_0 + \gamma_1 ICOL_i + (\gamma_2 - b)ECOST_j + \gamma_3 ICOL_i ECOST_j + \varepsilon_{ij}$$

and, with the maintained distributional assumptions,

$$(8)\quad \Pr(T_{ij} = 1 \,|\, \mathbf{X}) = \Lambda[\beta_0 + \beta_I ICOL_i + \beta_E ECOST_j + \beta_{IE} ICOL_i ECOST_j],$$

where $\hat{\beta}_0$, $\hat{\beta}_I$, $\hat{\beta}_E$, and $\hat{\beta}_{IE}$ are the implied estimates for the model parameters γ_0, γ_1, $(\gamma_2 - b)$, and γ_3, respectively.

B. *Interpreting the Model Coefficient Estimates*

Naturally, C_{ij} is expected to be increasing in $ICOL_i$, so that we predict $\hat{\beta}_I > 0$. This coefficient thus allows us to test whether a candidate measure of *ICOL* helps explain heterogeneity *among* individuals in total costs of lying. Moreover, specification (6) allows for two channels through which heterogeneity in total preferences for truthfulness *within* individuals enters. First, it seems reasonable to postulate that C_{ij} is increasing in *ECOST* so that individuals associate higher monetary stakes with a stronger preference to tell the truth. Given that the parameters b and γ_2 are not identified by the present approach, $\hat{\beta}_E < 0$ tells us only that C_{ij} is not increasing in *ECOST* at a rate greater than marginal utility b. Second, heterogeneity of preferences within individuals for truthfulness can unambiguously be detected by considering the interaction term β_{IE}. Under the discrete-choice model's assumptions, a significant interaction term provides further evidence against the notion, posited by

the type-based model, that there are two *fixed* types.[15] Specifically, a positive interaction term arises if the intrinsic costs of lying are more important in determining total preferences for truthfulness when the stakes (*ECOST*) are higher. An equivalent interpretation (useful in settings where an agent can choose the size of the lie) is that individuals with stronger intrinsic costs of lying perceive "larger" lies, which yield larger economic benefits, as less attractive and will, thus, tell "smaller" lies. A negative interaction term instead arises if the source of the intrinsic costs of lying is relatively less influential at higher stakes. If the coefficient on the interaction is zero, the economic costs of truthtelling are perceived identically by all agents, regardless of the strength of their intrinsic costs of lying, $ICOL_i$. In that case, all agents' utilities would react identically to changes in the economic costs, even though heterogeneous $ICOL_i$ would imply that some would report the truth while others would lie at a given *ECOST*.[16]

C. *Results*

We first consider three possible sources of intrinsic costs of lying, $ICOL_i$, for which survey measures are available to us, and we then discuss other possible drivers of behavior. The descriptive statistics for *EXTDECEIT*, *SELFDECEIT*, and *PV* shown in Table 4 indicate that there is wide variation in these three variables, suggesting that they could potentially explain the observed variation in truthtelling behavior. In Table 3, we test whether this is the case. We allow each possible source to affect behavior both separably from economic costs and jointly by way of an interaction.

First, participants may have developed an interest in impressing the experimenter by appearing honest and nongreedy (e.g., Fischbacher and Heusi 2008); this would act like a preference for truthfulness. Given the design of the experiment, in which we took great care to make the responses anonymous, this is very unlikely to have occurred. Indeed, *EXTDECEIT* is not significant in any of the regressions, and neither is the interaction term with *ECOST*.

[15] Alternatively, if the true utility function has a separable form but the assumption of weak exogeneity of the error term ε_{ij} does not hold, then the interaction term in the model may serve as an instrument to correct for correlation between the explanatory variables ($ICOL_i$ and $ECOST_j$) and the error term ε_{ij}. The interaction term can serve as an instrument because it arises as one of the terms in the second-order Taylor-series expansion of the random utility function (with violated weak exogeneity). It captures jointly the effects of both explanatory variables. One anonymous referee provided an example in which 90 percent of individuals make deterministic decisions based on an additively separable utility function, i.e., they tell the truth when $C_i = ICOL_i$ is greater than *ECOST*, and they lie when $ICOL_i$ is smaller than *ECOST*. Ten percent of individuals make decision errors; that is, they tell the truth although *ECOST* is greater than $ICOL_i$, and they lie even when $ICOL_i$ is greater than *ECOST*. In simulated data, the referee showed that a logit regression (which is a misspecified model under the assumptions made) may yield a significant interaction term on *ECOST* and intrinsic costs of lying, even in this setting. In the referee's example, the error term is not independent of the observable variables. Thus, the interaction term becomes significant in this setting because it is an instrumental variable for an omitted variable. We conclude that, even if the true utility function is separable, researchers may well want to use a nonseparable reduced form because this specification is robust to the violation of the assumption of weak exogeneity of the error term, such as occurs in decision errors of the form proposed by the referee.

[16] When testing for nonseparability (that is, for the significance of the interaction term), we consider coefficients, rather than marginal effects, from the logit regressions. Recall that marginal effects in a logit regression are given by $\Lambda(\boldsymbol{\beta}'\mathbf{X})(1 - \Lambda(\boldsymbol{\beta}'\mathbf{X}))\boldsymbol{\beta}$, where $\Lambda(\cdot)$ is the logistic cumulative distribution function giving the initial probability of truthfulness. Those with high (low) *ICOL* have high (low) initial probabilities of truthfulness. Thus, the highest marginal effects of *ECOST* on behavior are likely to be found in the middle range, and smaller marginal effects are likely to be found among those with high intrinsic costs. Analyzing coefficients instead allows us to consider the hypothetical case of participants who would display identical initial probabilities of reporting the truth.

TABLE 4—DESCRIPTIVE STATISTICS OF IMPORTANT EXPLANATORY VARIABLES

	Mean	Median	SD	Min.	Max.
Altruistic concerns (35*HURTS*)	0.74	1.00	1.07	−2.00	2.00
Tendency towards impression management (*EXTDECEIT*)	0.49	0.50	0.13	0.16	0.81
Tendency towards self-deception (*SELFDECEIT*)	0.64	0.63	0.12	0.24	1.00
Protected values (*PV*)	3.82	3.78	1.03	0.00	6.00

Notes: This table presents descriptive statistics for our measure of altruistic concerns and for three candidate measures of intrinsic costs of lying. $N = 261$.

Second, it is possible that participants deceived themselves by making the "right" choices. However, *SELFDECEIT* is also not significant in any of the regressions, and neither is the interaction term with *ECOST*.

Third, we consider the possibility that moral values were a source of the intrinsic costs of lying. While many moral concepts are potentially relevant, we focus on protected values (*PV*). The literature that has developed the theory of these values emphasizes that protected values are nonconsequentialist and induce a resistance to engaging in actions that would violate moral values, reducing the attractiveness of any financial gains obtained through such actions.[17] That is, the economic costs of truthfulness matter less to those who hold stronger protected values of truthfulness; those people are trade-off resistant. This idea naturally translates into a functional form for C_i that is nonseparable into intrinsic (moral) costs of lying and economic costs of truthfulness.

Column 3 of Table 3 shows that *PV* of truthfulness was a highly significant predictor of behavior in the experiment. A one-point increase in *PV* was associated with a 17.4 percent increase in the probability of truthtelling, holding the other variables at their means. In column 4, we obtain a positive, significant coefficient on the interaction term between *PV* and *ECOST*. This is evidence that, conditional on the correctness of the discrete choice (logit) model's specification, the data are consistent with nonseparability of the economic incentives and this measure of intrinsic costs of lying. That is, the data confirm that there is heterogeneity within individuals' total costs of lying, again inconsistent with the type-based model's assumption.

Note that $\hat{\beta}_E + \hat{\beta}_{IE} PV_i$ is negative even when evaluated at $PV = 6$. Thus, in the cross-section of participants, the presence of a strong protected value of truthfulness lessened, but did not eliminate, the relevance of the economic costs associated with the earnings management decision. With *PV* in the regression, the significance of the demographic controls vanishes. It is also noteworthy that, as soon as we include the interaction term with *PV*, 35*HURTS* is no longer significant.[18] Finally, as shown

[17] See, for example, Baron and Spranca (1997); Tetlock et al. (2000); and Tanner, Medin, and Iliev (2008). The source of protected values is modeled by Bénabou and Tirole (2011) as a need of agents to invest in their identity. For the strongest form of *PV*, "sacred" values and taboos, see in particular Section V of their paper.

[18] This is as expected, given the setup of this experiment, and it suggests that the significance of 35*HURTS* in the earlier regressions stems from the fact that this variable (as well as the underlying social-preferences intensity of the individual) is correlated with intrinsic costs of lying. 35*HURTS* has a positive correlation with *PV* of 0.34. Within our experimental setup, we are unable to address any possible fundamental relationship between protected values and altruistic concerns, so we leave this to future research.

in column 5, we also find that our results continue to hold in the subsample without the free-truth situation.

We emphasize that, despite these findings, one cannot conclude that *PV* has a stronger claim to organizing the data than plausible alternatives. For example, participants may be driven by nonconsequentialist preferences that attach expressive utility to low-stakes acts or decisions that substantiate or confirm personal identity. This expressive-preferences concept was developed in the political science literature to explain why citizens vote despite an apparent lack of economic incentive (Buchanan 1954; Tullock 1971). Expressive preferences have been experimentally documented to play a role in hypothetical choice situations, for example, by Feddersen, Gailmard, and Sandroni (2009). While, in the formulation of these authors, expressive preferences (*EP*) enter utility separably from economic costs, it is conceivable that *EP* also interact with incentives so that nonseparability arises. To the extent that the *PV* survey implicitly measures *EP*, and conditional on the correctness of the statistical model's specification, the results on the interaction term imply that *EP* may, in the range of economic stakes considered in this experiment, become more important in creating a difference in the perceived attractiveness of truthfulness and lying as the stakes increase.[19] Overall, the evidence available from this experiment does not allow us to discriminate definitively between *EP* and *PV* as possible sources of heterogeneity in agents' lying behaviors. Additional factors relevant in the real world are also not addressed here. For example, intrinsic costs of lying may also be due to an internal reward mechanism for truthfulness that is activated when individuals are, for example, asked to recall the Ten Commandments or to sign an honor code (e.g., Mazar, Amir, and Ariely 2008).

D. *Further Results and Robustness*

Conceivably, participants may have worried about the wealth of the experimenters, which would show up in systematic variation in their choices in the effort task. But there is no observable relationship between the participants' levels of effort in that task and their *PV*, 35*HURTS*, or socially acceptable responding. This finding also confirms that the experimental design did not simply produce the same pattern of results in the truthtelling and the effort tasks. Moreover, our results are robust to controlling for investment experience, and to variations in samples and estimation methods. All these additional results are available upon request.

V. Conclusion

In this study, we examined individuals who were exposed to a simple but realistic trade-off: they could tell the truth and suffer economic costs of truthfulness, or they could lie and potentially incur intrinsic costs of lying. In our setting, there was no strategic incentive to tell the truth; the participants had no counterparty, no notion of a repeated game, no legal obligation, and no risk of being punished.

[19] Conceptually, *EP* may also be related to tendencies towards self-deception, though the previous results suggest that a standard measure of such tendencies does not explain behavior in this experiment.

The experimental results unambiguously reject a type-based model. That is, the results refute Hypothesis TYP, that there exist only "the Ethical" (who care so much about the rightfulness of the process that they always tell the truth) and "the Economic" (who care only about their material payoffs and, thus, always lie when profitable). Instead, this paper supports Hypothesis HET, reflecting Gneezy's (2005) conjecture of continuous heterogeneity of preferences for truthfulness: people balance process against consequences in a range of different ways. Moreover, we provide evidence that preferences for truthfulness are nonseparable in intrinsic preferences and economic incentives. In sum, our findings point to heterogeneity, both among and within individuals, in their preferences for truthfulness.

This experiment cannot definitively identify the ultimate source(s) of the intrinsic costs of lying. Nor can it state whether the suggested preferences for truthfulness would also be at work at much higher stakes (as the protected-values explanation implies) or whether the validity of the results is limited to relatively low-stakes settings (as the expressive-preferences explanation suggests). Future research may be fruitfully conducted to answer these important questions.

To the extent that preferences for truthfulness apply in a wide range of settings, the results obtained in this study have implications regarding the effectiveness of methods to screen agents for their preferences for truthfulness, as well as implications for the optimal setting of incentive contracts.

REFERENCES

Ariely, Dan, Anat Bracha, and Stephan Meier. 2009. "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially." *American Economic Review* 99 (1): 544–55.

Baiman, Stanley, and Barry L. Lewis. 1989. "An Experiment Testing the Behavioral Equivalence of Strategically Equivalent Employment Contracts." *Journal of Accounting Research* 27 (1): 1–20.

Baron, Jonathan, and Mark Spranca. 1997. "Protected Values." *Organizational Behavior and Human Decision Processes* 70 (1): 1–16.

Bénabou, Roland, and Jean Tirole. 2011. "Identity, Morals, and Taboos: Beliefs as Assets." *Quarterly Journal of Economics* 126 (2): 805–55.

Bergstresser, Daniel, and Thomas Philippon. 2006. "CEO Incentives and Earnings Management." *Journal of Financial Economics* 80 (3): 511–29.

Bhide, Amar, and Howard H. Stevenson. 1990. "Why Be Honest if Honesty Doesn't Pay?" *Harvard Business Review* 68 (5): 121–29.

Bowles, Samuel, and Sandra Polania-Reyes. 2012. "Economic Incentives and Social Preferences: Substitutes or Complements?" *Journal of Economic Literature* 50 (2): 368–425.

Buchanan, J. M. 1954. "Individual Choice in Voting and the Market." *Journal of Political Economy* 62: 334–43.

Charness, Gary, and Martin Dufwenberg. 2006. "Promises and Partnership." *Econometrica* 74 (6): 1579–601.

Evans, John H. III, R. Lynn Hannan, Ranjani Krishnan, and Donald V. Moser. 2001. "Honesty in Managerial Reporting." *Accounting Review* 76 (4): 537–59.

Feddersen, Timothy, Sean Gailmard, and Alvaro Sandroni. 2009. "Moral Bias in Large Elections: Theory and Experimental Evidence." *American Political Science Review* 103 (2): 175–92.

Fehr, Ernst, and Urs Fischbacher. 2002. "Why Social Preferences Matter—The Impact of Non-selfish Motives on Competition, Cooperation and Incentives." *Economic Journal* 112 (478): C1–33.

Fehr, Ernst, and Urs Fischbacher. 2003. "The Nature of Human Altruism." *Nature* 425: 785–91.

Fehr, Ernst, and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics* 114 (3): 817–68.

Fischbacher, Urs, and Franziska Heusi. 2008. "Lies in Disguise: An Experimental Study on Cheating." Unpublished.

Gibson, Rajna, Carmen Tanner, and Alexander F. Wagner. 2013. "Preferences for Truthfulness: Heterogeneity among and within Individuals: Dataset." *American Economic Review*. http://dx.doi.org/10.125/aer.103.532.

Gneezy, Uri. 2005. "Deception: The Role of Consequences." *American Economic Review* 95 (1): 384–94.

Gneezy, Uri, Alex Imas, and Kristóf Madarász. 2012. "Conscience Accounting: Emotional Dynamics and Social Behavior." Unpublished.

Grover, Steven L. 2005. "The Truth, the Whole Truth and Nothing but the Truth: The Causes and Management of Workplace Lying." *Academy of Management Executive* 19 (2): 148–57.

Healy, Paul M., and James M. Wahlen. 1999. "A Review of the Earnings Management Literature and its Implications for Standard Setting." *Accounting Horizons* 13: 365–83.

Hurkens, Sjaak, and Navin Kartik. 2009. "Would I Lie to You? On Social Preferences and Lying Aversion." *Experimental Economics* 12 (2): 180–92.

King, Gary. 1998. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference.* Ann Arbor: University of Michigan Press.

Koford, Kenneth, and Mark Penno. 1992. "Accounting, Principal-Agent Theory, and Self-Interested Behavior." In *Ethics and Agency Theory: An Introduction,* edited by Norman E. Bowie and R. Edward Freeman, 127–42. New York: Oxford University Press.

Mazar, Nina, On Amir, and Dan Ariely. 2008. "The Dishonesty of Honest People: A Theory of Self-Concept Maintenance." *Journal of Marketing Research* 45 (6): 633–44.

Musch, Jochen, Robbi Brockhaus, and Arndt Bröder. 2002. "Ein Inventar zur Erfassung von zwei Faktoren sozialer Erwünschtheit." *Diagnostica* 48 (3): 121–29.

Nelson, Mark W., John A. Elliott, and Robin L. Tarpley. 2003. "How are Earnings Managed? Examples from Auditors." *Accounting Horizons* 17S: 17–35.

Paulhus, Delroy. 1984. "Two-Component Models of Socially Desirable Responding." *Journal of Personality and Social Psychology* 46 (3): 598–609.

Rabin, Matthew. 1995. "Moral Preferences, Moral Constraints, and Self-Serving Biases." Unpublished.

Rode, Julian. 2010. "Truth and Trust in Communication: Experiments on the Effect of a Competitive Context." *Games and Economic Behavior* 68 (1): 325–38.

Sánchez-Pagés, Santiago, and Marc Vorsatz. 2007. "An Experimental Study of Truth-Telling in a Sender-Receiver Game." *Games and Economic Behavior* 61 (1): 86–112.

Sutter, Matthias. 2009. "Deception through Telling the Truth?! Experimental Evidence from Individuals and Teams." *Economic Journal* 119 (534): 47–60.

Tanner, Carmen, Douglas L. Medin, and Rumen Iliev. 2008. "Influence of Deontological Versus Consequentialist Orientations on Act Choices and Framing Effects: When Principles are More Important than Consequences." *European Journal of Social Psychology* 38 (5): 757–69.

Tanner, Carmen, Bettina Ryf, and Martin Hanselmann. 2009. "Geschützte Werte Skala: Konstruktion und erste Validierung eines Messinstrumentes (Protected Values Measure: Construction and First Validation of an Instrument to Assess Protected Values)." *Diagnostica* 55 (3): 174–83.

Tetlock, Philip E., Orie V. Kristel, S. Beth Elson, Melanie C. Green, and Jennifer S. Lerner. 2000. "The Psychology of the Unthinkable: Taboo Trade-Offs, Forbidden Base Rates, and Heretical Counterfactuals." *Journal of Personality and Social Psychology* 78 (5): 853–70.

Treviño, Linda K., and Michael E. Brown. 2004. "Managing to be Ethical: Debunking Five Business Ethics Myths." *Academy of Management Executive* 18 (2): 69–81.

Tullock, Gordon. 1971. "The Charity of the Uncharitable." *Economic Inquiry* 9 (4): 379–92.

Wooldridge, Jeffrey M. 2006. *Introductory Econometrics: A Modern Approach.* Mason, OH: Thomson Higher Education.

Zak, Paul J., ed. 2008. *Moral Markets: The Critical Role of Values in the Economy.*, NJ: Princeton University Press.