

Swiss Finance Institute  
Research Paper Series N°08 – 48

# Ethics Can Trump Self-Interest: Experimental Evidence on the Protected Value of Truthfulness as a Motivator for Truth-Telling

Rajna GIBSON  
University of Geneva and Swiss Finance Institute

Carmen TANNER  
University of Zurich

Alexander F. WAGNER  
University of Zurich and Swiss Finance Institute

swiss:finance:institute

 FINRISK  
National Centre of Competence in Research  
Financial Valuation and Risk Management

Established at the initiative of the Swiss Bankers' Association, the Swiss Finance Institute is a private foundation funded by the Swiss banks and SWX. It merges 3 existing foundations: the International Center FAME, the Swiss Banking School and the Stiftung "Banking and Finance" in Zurich. With its university partners, the Swiss Finance Institute pursues the objective of forming a competence center in banking and finance commensurate to the importance of the Swiss financial center. It will be active in research, doctoral training and executive education while also proposing activities fostering interactions between academia and the industry. The Swiss Finance Institute supports and promotes promising research projects in selected subject areas. It develops its activity in complete symbiosis with the NCCR FinRisk.

The National Centre of Competence in Research "Financial Valuation and Risk Management" (FinRisk) was launched in 2001 by the Swiss National Science Foundation (SNSF). FinRisk constitutes an academic forum that fosters cutting-edge finance research, education of highly qualified finance specialists at the doctoral level and knowledge transfer between finance academics and practitioners. It is managed from the University of Zurich and includes various academic institutions from Geneva, Lausanne, Lugano, St.Gallen and Zurich. For more information see [www.nccr-finrisk.ch](http://www.nccr-finrisk.ch).

This paper can be downloaded without charge from the Swiss Finance Institute Research Paper Series hosted on the Social Science Research Network electronic library at:

<http://ssrn.com/abstract=1323283>

**Ethics *Can* Trump Self-Interest:  
Experimental Evidence on the Protected Value of Truthfulness  
as a Motivator for Truth-Telling <sup>♦</sup>**

Rajna Gibson <sup>\*</sup>, Carmen Tanner <sup>\*\*</sup>, and Alexander F. Wagner <sup>\*\*\*</sup>

October 20, 2010

**Abstract**

There are people who are motivated by the non-selfish, non-strategic, and non-consequentialist “protected value” of truthfulness. We conduct an experiment directly assessing this phenomenon. We find that people differ substantially in their truthfulness, with a large minority powerfully inclined to honesty. Those who consider truthfulness to be a protected value are willing to make significant monetary sacrifices. They are less concerned than others with changes in truth-telling costs and with social norms. These findings have implications for agent selection and incentive design.

Keywords: Ethics, incentives, protected values, social norms, earnings management, truthfulness, honesty

JEL-Codes: A13, C91, G30, M14

---

<sup>♦</sup> 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 for programming the first version of the code for the experiments. We thank the Editor and two anonymous referees for excellent remarks that greatly improved the paper. We also thank Peter Bossaerts, Chiddi Chidambaran, Alain Cohn, François Degeorge, Ernst Fehr, Gerlinde Fellner, Laurent Frésard, Susanne Neckermann, Meir Statman, Mei Wang, and Rainer Winkelmann for fruitful discussions and comments. Useful comments were provided by seminar participants at the European Finance Association meetings, the International Conference of 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.

<sup>\*</sup> Swiss Finance Institute Professor of Finance, University of Geneva. Corresponding author: [Rajna.Gibson@unige.ch](mailto:Rajna.Gibson@unige.ch), l'Université de Genève, 40 Bd Pont d'Arve, CH-1211 Genève 4, Switzerland, Tel: +41 (022) 379 8983

<sup>\*\*</sup> Professor of Psychology, Department of Psychology, University of Zurich

<sup>\*\*\*</sup> Swiss Finance Institute Assistant Professor of Finance, University of Zurich

# 1. Introduction

Standard economic models of decision making, which emphasize the role of consequences (usually expressed in utility terms) in determining agents' actions, seem to explain a grim inexorability to all economic systems. These conventional models adopt the view of human nature that people are self-interested and behave dishonestly for cogent reasons. When behaving as expected, as pure utility maximizers, agents prioritize the outcomes of their actions and only forgo materially beneficial lying if strategic or reputational considerations arise. Some researchers, such as Amar Bhidé and Howard H. Stevenson (1990), assert that these reputational forces are weaker than supposed, implying that honesty simply does not seem to pay.

Examples of disastrous dishonesty based on such self-interest abound in the corporate world. In the recent financial crisis, the economic effects of regulatory failure, of a deteriorating macro-economy, and of inadequate models were certainly augmented by deliberate deception. In fact, the incentives to lie and even defraud may have strengthened, in part as a response to this financial crisis.

Yet, unnoticed in the tumult may be the bedrock that we seek. Even casual observation suggests that truthfulness does seem to prosper in society. Whistleblowers (even in companies that establish opportunities for anonymity) often jeopardize their careers and friendships when they truthfully reveal the wrongdoing of their companies. Some CEOs are regarded as particularly virtuous (Linda K. Treviño and Michael E. Brown 2004). Numerous journalists have risked (and even lost) their lives for reporting the truth about political repression, economic crimes, and human rights violations.

Several positive theories on ethical or moral behavior have recently been developed; these recognize that people do not necessarily maximize utility according to the material consequences of their actions.<sup>1</sup> Furthermore, a number of studies have documented experimentally that individuals' choices to tell the truth or lie deviate from the predictions of utilitarian economic models.<sup>2</sup> Puzzled by the extent to which people tell the truth even when it

---

<sup>1</sup> See, for example, the work by Thomas H. Noe and Michael J. Rebelló (1994), Roland Benabou and Jean Tirole (2010), Michael C. Jensen (2007), Brian Mittendorf (2007), Bruce Carlin and Simon Gervais (2008), and Paul Fischer and Steven Huddart (2008).

<sup>2</sup> Below we discuss the work of Stanley Baiman and Barry L. Lewis (1989), John H. Evans, R. Lynn Hannan, Ranjani Krishnan, and Donald V. Moser (2001), Urs Fischbacher and Franziska Heusi (2008), Uri Gneezy (2005), Sjaak Hurkens and Navin Kartik (2009), Nina Mazar, On Amir, and Dan Ariely (2008), Julian Rode (2010), Santiago Sánchez-Pagés and Marc Vorsatz (2007), and Matthias Sutter (2009). See the edited volume by Paul J. Zak (2008) for numerous additional examples.

is costly to do so, researchers, such as Gneezy (2005), have often ascribed the observed behavior to a latent preference for truthfulness.

In this study, we conduct an experiment to test the conjecture that there exist individuals who base their economic decisions on moral commitments rather than on consequences for themselves and/or for others. These people would consider truthfulness to be a protected value and would, therefore, treat truth-telling as morally mandatory. Due to this commitment, they would be expected to be less influenced by economic and social incentives (such as social approval or disapproval).

For a conceptual framework for our experiment, we draw upon recent work in psychology on the notion of “protected values” as a powerful principle for organizing an understanding of decision making.<sup>3</sup> This concept of protected values, those not available for trade-off with economic values, encompasses moral imperatives that drive actions independently of material considerations. The most extreme version of the protected values concept would predict rejection of any compromise at all, but this rarely occurs in reality. A less extreme version would characterize moral motivation as one important consideration among several, having, however, greater force with people who hold it as a protected value. These people also would show more trade-off resistance and less willingness to compromise than people without protected values.<sup>4</sup> We provide a preference specification that captures this idea.

This protected-values concept allows us to explore a foundation of honesty that is both non-selfish and non-consequentialist. With this, we add to a long-standing literature that considers the role of preferences that depend on process and/or on consequences for others. For example, in a seminal paper, Matthew Rabin (1993) demonstrated how fairness considerations can explain why people are willing to reward or punish others even though this requires a sacrifice of their own well-being. Gneezy (2005), in addressing people’s propensity to lie, emphasized the joint relevance of process-dependent preferences and of consequences to oneself and to others. In a cheap-talk sender-receiver game, Gneezy found that many subjects told the truth in a setting where standard economic theory predicts that lying would be the dominant action. He concluded that “people not only care about their own gain from lying; they are also sensitive to the harm their lie may cause to the other side” (page 391). Gneezy suggested adjusting the standard economic model by employing non-consequential preferences.

---

<sup>3</sup> See, for example, Jonathan Baron and Mark Spranca (1997), Philip E. Tetlock, Ori V. Kristel, S. Beth Elson, Melanie C. Green, and Jennifer S. Lerner (2000), Tetlock (2003), Scott Atran, Robert Axelrod, and Richard Davis (2007), and Carmen Tanner, Douglas L. Medin, and Rumen Iliev (2008).

<sup>4</sup> Recent findings, for instance, showed that people with protected values often have no problems accepting trade-offs when the amount or probability of violation of protected values is small (for example, Baron and Sarah Leshner (2000)), while still preserving strong opinions about their protected values.

Similar fundamental results on truth-telling have been obtained in other studies (Evans et al. 2001; Sánchez-Pagés and Vorsatz 2007). Only a few studies, such as Baiman and Lewis (1989), have found that people will lie even for just a tiny monetary payoff. However, there is disagreement regarding the conclusions that can be reached from these sender-receiver games. First, 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). Second, Rode (2010) showed the importance of the environment-driven beliefs of decision makers who receive information from better-informed advisors. He found that decision makers were significantly less trusting in a competitive context than in a cooperative context. Finally, Hurkens and Kartik (2009) showed that Gneezy's (2005) empirical observations were possibly "entirely due to people's social preferences over different allocations, and not because of how lying aversion varies with allocations" (p. 181). Hurkens and Kartik showed that Gneezy's results were also consistent with an alternative hypothesis: that, conditional on preferring the outcome of lying to the outcome of truth-telling, a person may be sensitive neither to her own (monetary) gain from lying, nor to the (monetary) harm she causes the other side. We interpret these findings as a caution against accepting the findings from sender-receiver games as evidence either for or against an inherent lying aversion.

Gneezy's study, among others, relied on aggregate behavior patterns to infer the possible existence of preferences for truth-telling. Such aggregate evidence has also been employed outside the realm of sender-receiver games. For example, Fischbacher and Heusi (2008) found that some experimental participants eschewed making false claims that would have earned them more money than incomplete lies. Therefore, Fischbacher and Heusi conjectured that participants cared not only about their monetary gains but also about being perceived as honest and non-greedy. Supporting the concept of an internal reward mechanism that can increase truthfulness, Mazar, Amir, and Ariely (2008) conducted an experiment in which merely asking participants to recall the Ten Commandments or to sign an honor code caused an increase in the honest behavior of the group.<sup>5</sup>

We offer two contributions to the existing literature. First, we observe individuals' behaviors and values and explore how behavior varies among and within individuals as the economic costs of truthfulness and the social norms (that is, the approval or disapproval of particular behaviors) regarding truthfulness vary. In so doing, we provide stronger evidence regarding the heterogeneity of preferences for truthfulness than existing works. Second, our

---

<sup>5</sup> Altruism and reciprocity are two other facets of individual behavior that require going beyond the universal self-interest model. See, for example, Ernst Fehr and Fischbacher (2002) and Fehr and Fischbacher (2003) for surveys of some important recent research on these concepts.

simpler experimental setup, with only a decision-making situation, allows us to identify directly the motivations for truth-telling.

We conduct our analysis in the context of an experiment on accounting earnings management, which we will refer to as simply “earnings management.”<sup>6</sup> We choose this specific framework for several reasons. First, accounting earnings management has received widespread academic and practical attention lately. Second, it bears significant economic externalities. Billions of dollars in stockholder wealth and the fates of economic empires ride on earnings reports.<sup>7</sup> Third, it illuminates a salient potential conflict: Management’s variable compensation is frequently tied to stock price performance, which in turn often hinges on earnings announcements. Indeed, survey evidence (Ernst & Young 2009) indicates that a tough economic climate fosters financial statement fraud.

In our laboratory experiment, we observe that, in a situation where the standard utilitarian model clearly predicts that everybody will manage earnings, a significant minority of participants choose not to do so, thus forgoing a larger variable compensation. This is consistent with previous studies on costly truth-telling.

We then provide evidence that there exist individuals who consider truthfulness to be a protected value, and we establish the causal relationship of the value an individual places on truthfulness to his economic decision-making. Specifically, we study how persons who assign different values to truthfulness weigh that against the monetary benefits of lying. We also study how those people respond to social norms (Robert B. Cialdini, Raymond R. Reno, and Carl A. Kallgren 1990) that substitute for market pressure.

We first show that people’s observed economic behaviors can be explained by a measure of their regard for truthfulness, even after controlling for such factors as demographics, social norms, and the costs of truthfulness. This constitutes direct evidence of people’s different

---

<sup>6</sup> 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” (Paul M. Healy and James M. 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” (Steven L. Grover 2005) or as a “deliberate false statement that is either intended to deceive others or foreseen to be likely to deceive others” (Randi L. Sims 2000). As John R. Graham, Campbell R. Harvey, and Shiva Rajgopal (2005) point out, since the passage of the Sarbanes-Oxley Act, firms seem to be more actively pursuing real earnings management (for instance, by postponing an R&D investment or by speeding up sales by implementing marketing campaigns before the end of the fiscal year). Real earnings management is not only more difficult to identify and to estimate, but it is also less directly related to truth-telling and hence less relevant to our question.

<sup>7</sup> Of course, shareholders with a short-term orientation may actually prefer for a manager to engage in earnings management. Also, competition for funds may contribute to earnings manipulation, but such competition may ultimately be welfare-enhancing (Andrei Shleifer 2004). Addressing the conflict among shareholders regarding earnings management, and modelling competitive pressures among firms are tasks beyond the scope of this paper, and we design the experiment to exclude these facets.

levels of commitment to truthfulness and of the consequences of those differences on economic decision-making.

Several finer results of our experiment support this causal relationship between commitments to truthfulness and economic decision-making. The protected-values framework implies that individuals with stronger stated values for truthfulness will react less sensitively to monetary incentives for lying. As we manipulate the costs of truthfulness, we find that, consistent with this prediction, those with strong protected values are much more resistant to the influence of changing costs.

Additionally, we manipulate the social norms by informing participants of society's supposed acceptance or disapproval of earnings management. It has long been established in the literature that norms are a powerful driving force in human behavior. Our novel contribution is to show that the behavior of individuals with the strongest protected values tends to be the least sensitive to social norms. The evidence here is suggestive and not quite as statistically significant as that regarding individuals' resistance to direct economic costs.

Importantly, we obtain results indicating, among those with strong protected values, greater imperviousness to varying economic costs and social norms, as we control for the fact that these participants are initially more likely to report the truth than those with weak protected values.

We also find that those participants with strong protected values for truthfulness do not engage more or less industriously than those with weak protected values in the effort-based task of making simple calculations. This serves as an indication that it is unlikely that spurious causality or an experimenter-demand effect is driving our findings in the truth-telling task.

Experimental simulation of corporate decision-making raises the question of external validity. Addressing this issue, several studies have shown that the levels of payments received by participants do not have a dramatic effect on their behavior if the subjects are paid proportionately to the opportunity cost of their time.<sup>8</sup> In addition, the available scientific evidence and the conclusions of evolutionary psychology suggest that the findings from our experiment have not only local validity but also relevance to similar situations in a wide range of settings.<sup>9</sup> Finally, we do not find that students of economics and students of psychology

---

<sup>8</sup> See, for example, Douglas D. Davis and Charles A. Holt (1992), Elizabeth Hoffman, Kevin A. McCabe, and Vernon Smith (1996), and Robert Forsythe, Joel L. Horowitz, N. E. Savin, and Martin Sefton (1991).

<sup>9</sup> In particular, this research in evolutionary psychology shows that similar ecologically relevant problems are solved in a roughly similar way by most humans (controlling for their individual characteristics, such as the strengths of their values). Numerous studies find that the behavior of professional decision makers does not qualitatively differ from that exhibited by student subject groups, among them, Douglas V. DeJong, Forsythe, and Wilfred C. Uecker (1988), Douglas Dyer, John H. Kagel, and Dan Levin (1989), Orly Sade, Charles Schnitzlein, and Jaime F. Zender (2006), and Smith, Gerry L. Suchanek, and Arlington W. Williams (1988). See also Ann B.

behave significantly differently in our experiment, even though their educational backgrounds differ quite substantially. Despite these pieces of evidence, we cannot ascertain that student participants behave identically to corporate executives.

In sum, by considering individual-level data on values and behavior and by leaving aside strategic concerns, this experiment offers evidence that truthful reporting is, at least to some extent, based on a non-selfish and non-strategic consideration of truthfulness as a protected value in and of itself, independent of its consequences for the agent or for others.

The rest of this paper is organized as follows. Section 2 presents a stylized decision-making model that introduces the assumptions and the predictions of this experiment. Section 3 describes the experimental method and design. Section 4 discusses the experimental results. Section 5 offers concluding remarks.

## **2. Protected values as motivation for truth-telling**

We are interested in economic situations in which individuals have a choice between telling the truth and telling a lie, and in which the former decision involves an economic sacrifice. Accounting earnings management is an important example of this type of situation. We have envisioned a framework in which earnings management is legal (for example, within GAAP rules), but in which participants know that investors would be best served by the absence of earnings management.<sup>10</sup> Of course, in practice, earnings management may be in the interest of (especially short-term) shareholders. We, therefore, have constructed a situation in which earnings management is explicitly self-interested and unequivocally unethical.

The advantage of designing this experiment is that we can tease out any causal influence of protected values on truth-telling behavior, while controlling for a select number of other important influences. For example, this experiment allows us first to examine whether and how preferences for truthfulness determine the choices of individuals. (Future work may consider much more complex and possibly more realistic group decisions.) We consider the

---

Gillette, Noe, and Rebello (2008). If anything, real CEOs have been found to be more trusting (as well as trustworthy) and, thus, less conforming to the economic model of self-interest than students (Ernst Fehr and John A. List 2004).

<sup>10</sup> Mark W. Nelson, John A. Elliott, and Robin L. Tarpley (2003) provided many examples collected from auditors. One classic example (p. 24) is the situation in which, upon closing the books, the managers found that the company's earnings were short compared to market expectations. The managers then revisited their deferred tax asset valuation allowance. The anticipated benefit realization was revised to "more likely than not" for certain foreign tax credits, and a portion of the allowance was reversed and credited to income. The amount resulted in earnings per share equal to market expectations. There are also examples within organizations. For example, a local bank branch understated the allowance for loan losses such that current earnings and capital were overstated. Local management wanted to show a positive net income, and favorable results affected the determination of local management's compensation (p. 23).

participants' choices, in the form of CEO-style decisions regarding engaging in earnings management. In the decision-making problem, we focus exclusively on the managerial choice. We require the recipient (the market, played by the computer) to accept passively all financial statements. The significant advantage of this approach is that, due to the absence of strategic interactions, we can isolate, at least better than in the real world, the factors influencing individuals' choices, without keeping track of the participants' thoughts regarding the behavior of other players.

The concept of protected values (*PVs*) was originally developed outside of the utility-maximization paradigm. In order to organize the experimental findings, we present a stylized model, using notation familiar to economists, to capture the fact that individuals may care, though to different extents, about both the act of telling the truth and the monetary consequences of their actions.<sup>11</sup>

Assume that the preferences of manager  $i$  are represented by the global utility function

$$V_i(y(T), T | PV_i) = u_i(y(T) | PV_i) - c(PV_i)(1 - T) \quad (1)$$

where  $u_i(\cdot)$  is an increasing and concave utility function,  $y$  is a flow of income or consumption that depends on whether the manager tells the truth,  $c(PV_i)$  captures the moral costs of lying,  $PV_i$  is the manager's protected value, and  $T=1$  when the manager chooses to report the truth, while  $T=0$  otherwise. Conditioning global utility on  $PV_i$  emphasizes that we are introducing a formulation that recognizes the possibility that manager  $i$ 's protected values may lead him to report the truth regardless of the consequences stemming from such an act.<sup>12</sup>

Note that this formulation allows protected values to be a factor in the global utility function in two possible ways: as a term that is independent of the flow of income and as part of the valuation of an income stream. The first effect of protected values is that they increase the moral costs of lying per se,  $c(PV_i)$ . We will denote these moral costs by  $c_i$ .

For the second effect of protected values, we consider the flow of income a manager receives. We refer to accounting earnings management as the action which amounts to lying, and to a lack of accounting earnings management as the action which amounts to truthfulness. When a manager lies, he receives a certain income  $x$ . There are economic costs of truthfulness, with  $0 \leq a \leq 1$  a parameter measuring that economic cost. For example, a manager expects the

<sup>11</sup> For a more elaborate model of commitment to truthfulness (and other values) as an investment in identity, see Benabou and Tirole (2010).

<sup>12</sup> We are interested only in those motivations that are rooted in protected values for individuals to tell the truth. In this experiment, we also control for other individual characteristics, such as a tendency towards self-deceit. Note that this reduced form treats truthfulness as a matter of preference. It would also be possible for there to be a constraint on behavior, with that constraint including a need to maintain a minimum level of truth-telling. 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.

stock price to drop if he does not manage earnings to match the market's expectations. If the compensation of the manager is tied to stock price performance, his failure to manage the earnings is expected to result in a loss of, at least, personal income for him.

Here, the assumption we make regarding the potential role of protected values is that the value an agent ascribes to monies depends on the origin of funds and on his commitment to truth-telling. Specifically, a dollar obtained by lying may be regarded as less valuable than a dollar obtained by telling the truth. We thus posit that

$$y = \begin{cases} x(1-a) & \text{if } T=1 \\ x(1-a) + ax(1-k(PV_i)) & \text{if } T=0 \end{cases} \quad (2)$$

In other words, when the agent lies, he obtains more money, but the portion of funds obtained by lying is partially discounted by  $k(PV_i) = k_i$  where the latter denotes a discounting parameter which is assumed to be increasing in  $PV$ , with  $0 \leq k_i \leq 1$ . (If  $k_i > 1$ , the agent would derive extra net utility for bearing economic costs through truth-telling, a possible, though unlikely scenario.)

An individual exhibits truthfulness when

$$u_i(x(1-a)) > u_i(x(1-ak_i)) - c_i. \quad (3)$$

Most traditional economic models assume that  $c_i = k_i = 0$ , that a manager only cares about material consequences and not about the ethical aspects of the situation; thus, protected values are considered irrelevant. A manager maximizing utility would then naturally choose  $T=0$ ; she would manage earnings, and this choice would be invariant to  $a$ . (At  $a=0$ , the manager would perceive no advantage or disadvantage to either telling the truth or lying.) In most economic models, she would only deviate from this behavior when there were opportunistic (often strategic) or altruistic reasons that would increase  $x$  through truthfulness. In our experiment, we explicitly exclude strategic and altruistic concerns by focusing only on the participants' actions in one-shot decisions without feedback.

The strictest form of an absolutely protected value of truthfulness would require that individuals who hold such a value not react at all to the cost of truth-telling. In psychology, this refers to the principle of quantity insensitivity or absolute trade-off resistance. In economic terms, this corresponds to lexicographic preferences, represented by  $c_i = \infty$ . An individual for whom truthfulness is an absolute value will choose  $T=1$ , and this choice will be invariant to the economic cost  $a$ . That is, an actual manager who is fully committed to truthfulness will not manage earnings, even if that costs him a bonus, decreases the value of any equity or stock options he might hold, and causes him to lose his job.

Between the extremes of purely opportunistic managers and managers with an absolute commitment to truthfulness, there is likely to be a group of individuals who trade off the benefits of truthfulness against the costs of monetary losses. That is, while they value the act of telling the truth, they also care about the consequences of their actions. As Gneezy (2005) points out, this position is descriptive even of St. Augustine, who is usually thought to have been categorically opposed to lying.<sup>13</sup>

The basic implication of Inequality (3) is that, with all other factors remaining the same, agents with higher  $PV$  are more likely to tell the truth. Note that truthfulness can, in this framework, only arise as optimal behavior if there is a moral cost of lying per se,  $c_i > 0$ , that is independent of the flow of income. We are mainly interested in whether there exist individuals with a moral commitment to truthfulness, who will manage earnings less than others at a given economic cost of telling the truth. To test for this, we confront our participants with a choice between managing earnings, by reporting inflated earnings to the financial market, and not managing earnings, by announcing lower but accurate earnings numbers. We carefully instruct the participants that the inflated earnings number could be announced legally. Thus, we give the opportunistic model the best opportunity. If we then observe that some people still report the truth, this would be suggestive, but not conclusive, evidence of a commitment to truthfulness. Therefore, we consider also some more nuanced predictions.

First, applying Inequality (3) to a cross-section of managers with a range of  $PVs$  suggests that the percentage of managers engaging in earnings management will decrease in  $a$ . In the real world, managers are indeed faced with substantial cross-sectional and time-series variation in the economic cost of truth-telling: The greater the extent to which the manager's pay is tied to stock price performance, the greater will be that manager's incentive to inflate earnings. That is,  $a$  varies. The natural implication is that the higher this direct economic cost of truthfulness is, the smaller will be the number of managers who maintain an overriding commitment to truthfulness.<sup>14</sup> By contrast, a types-based model, such as that of Kenneth Koford and Mark Penno (1992), would suggest that the fraction of the subject population telling the truth would remain constant across varying costs because pure opportunists would always lie and pure ethical types would always tell the truth. Note also that Hurkens and Kartik (2009)

---

<sup>13</sup> Specifically, St. Augustine wrote, "To me, however, it seems certain that every lie is a sin..." But he did distinguish among different types of lies. Indeed, the continuation of the citation reads, "...though it makes a great difference with what intention and on what subject one lies..." (cited in Gneezy 2005, p. 384).

<sup>14</sup> This conjecture would be consistent with findings by Daniel Bergstresser and Thomas 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.

demonstrated that Gneezy's (2005) data could be, in principle, consistent with such a types-based model. Our experiment allows us to test such a conjecture.

Second, we predict that those with stronger protected values will be more resistant to changes in the economic cost of truthfulness. This is because, if  $k_i > 0$  and increasing in  $PV$ , the difference in utility obtained from reporting the truth and from lying does not change as much for those with strong protected values. To test this prediction, the empirical analysis will need to consider the hypothetical case that the initial probabilities of truth-telling were the same for individuals with heterogeneous protected values. If  $c_i > 0$  but  $k_i = 0$ , the economic costs of truth-telling are perceived identically by all agents, no matter how strong their protected values are. Then, all agents would react identically to changes in the economic costs (but heterogeneous  $c_i > 0$  would imply that some would report the truth while others would lie at a given  $a$ ).

In our experiment, we also consider the influence of social norms. Much like protected values, this is a concept that does not play a role in standard economic theory; however, substantial evidence (especially from psychology) indicates that social norms guide human action in direct and meaningful ways (Robert B. Cialdini, Carl A. Kallgren, and Raymond R. Reno 1991, Deborah J. Terry and Michael A. Hogg 2001). Thus, society's general disapproval of earnings management or general acceptance of earnings management is likely to have an effect on behavior. In this experiment, we consider two types of norms: positive norms, which approve of lying, and negative norms, which disapprove of lying. One way to incorporate the role of norms into the preference specification is to posit that  $c_i = c(N, PV_i)$ , where  $N$  are social norms approving of lying, and  $c$  is decreasing in  $N$ . In this way,  $c_i$  now represents the moral and social costs of lying.

Society's general acceptance or society's disapproval of earnings management would then correspond to a larger or to a smaller  $N$ , respectively. Therefore, a positive (approving) norm should increase the fraction of managers who engage in earnings management; and a negative (disapproving) norm should decrease this fraction.

Our novel contribution here is to study whether some individuals react more strongly than others. Specifically, one might conjecture that the cross-derivative of  $c_i$  with respect to  $N$  and  $PV$  is negative. Intuitively, the moral and social costs of lying in the context of a positive norm (negative norm) would decrease (increase) less for those with strong protected values than for those with weak commitments to truthfulness. Thus, those with strong protected values would be the least likely to change their behavior simply because either type of social norm was in

place. Analogously to the case for the economic costs of truthfulness, we test whether, in the hypothetical case that the probabilities of truth-telling are the same for individuals with heterogeneous protected values, those with strong values react less sensitively to social norms than those with weak values.<sup>15</sup>

Finally, while we expect protected values of truthfulness to affect choice in a task that depends on  $T$ , those values should not matter in tasks for which this choice variable is neutral. In particular, we consider an observable effort choice task where effort  $e$  helps the agent to produce more output  $y(e)$ . Suppose effort cost is given by  $\psi(e)$ . A rational agent will maximize  $V_i(y(T), T | PV_i) = u_i(y(e, T)) - \psi(e)$ . Because effort will be observable in this setting,  $T$  will be irrelevant. Therefore, effort choice in this task will be independent of the strength of protected values of truthfulness.

### 3. Experimental method

#### PARTICIPANTS

A total of 261 subjects participated in this online experiment. We recruited our participants from undergraduate classes at the University of Zurich (Switzerland). 50% of the participants were economics and finance students, 40% psychology students, and 10% students from other fields. 42% were women and 58% were men (distributed across the fields). Their average age was 24.4 years. Participants received a fixed amount as payment for their participation and an additional variable amount as compensation determined by their performance. Anonymity was ensured.<sup>16</sup>

#### DESIGN

Each participant was placed in the situation of a CEO and confronted with two decision tasks: a truth-telling task and an effort task. These terms are used only here, for purposes of describing

---

<sup>15</sup> An alternative, but ultimately similar formulation, would posit that agents perceive or weigh the economic costs of truthfulness differently, depending on their protected values. Thus, the action-relevant, perceived economic costs of truthfulness for agent  $i$  are given by  $a_i = a(1 - k_i)$ , where  $k_i$  captures the notion of differential weighting of costs. If this function is increasing in  $PV$ , this implies the same comparative statics as in the text. In both formulations, the direct effect of norms will be strengthened further if  $k_i$  is also a decreasing function of  $N$ . And the greater imperviousness to social norms of individuals with high  $PVs$  is enhanced if  $k_i$  has a positive cross-derivative with respect to  $PV$  and  $N$ .

<sup>16</sup> In general, participants received the payment one week after the experiment. For this purpose, each participant had received, before the experiment, a code associated with his or her outcome. Based on this code, the experimenter prepared an envelope containing the earnings. Participants received the sealed envelopes by indicating their personal codes.

the experiment; the tasks were not labeled for the participants. In the truth-telling task, the participants were asked to report earnings per share in several situations offering the choice between accurate or inflated reporting. We manipulated the costliness of truthfulness, using five levels of truth-telling costs. Altogether, participants were given five choice tasks in both Phases 1 and 2. During the interval between Phases 1 and 2, we introduced a social norm manipulation that communicated whether earnings management was approved of or disapproved of by society. For this purpose, each participant was given a page that contained neutral, positive, or negative information about society's assessment of earnings management, thereby establishing one of the three conditions of social norms. The decision as to whether to tell the truth or to lie was the dependent variable in the truth-telling task. The independent, explanatory variables, the five levels of the cost of truthfulness in both Phases 1 and 2, were variations within the participants, while the social norms (neutral, positive, and negative) were variations among the participants.

In the effort task, the participants had to decide how much effort to expend. For that purpose, participants were confronted with simple calculation tasks, which consisted of adding and subtracting numbers. Participants were given five sets; in each of them they had to choose between doing one calculation and doing five calculations. By analogy with the previous task, we manipulated the level of the costs associated with effort. The decision as to whether to engage in low or high effort was the dependent variable in this task. Costs of effort, five levels from low to high, were variations within the participants.

Participants were randomly assigned to one of two orders of tasks: first the truth-telling and then the effort task, or vice versa; and to one of two orders of the five choice situations: from low to high costs or from high to low. We found no significant effect of these randomizations on any result. They were also randomly assigned to one of three conditions regarding the social norms manipulation: positive, negative, or neutral. We assessed protected values for truthfulness, together with some control variables, at the end of the experiment.

## **PROCEDURE**

All participants 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 the choices they made for each task.<sup>17</sup> The participants were then presented with the main parts of the experiment: 1) the truth-telling task, 2) the effort task, and 3) the measurement of

---

<sup>17</sup> At the time of the experiment, the exchange rate was about US \$1 = CHF 1.15.

protected values. The experiment ended with some final questions which mainly served as control variables, and then all the participants received their pay. As mentioned earlier, the order of the truth-telling and effort tasks was randomized. For simplicity, we describe the procedure for only one order of tasks. For both the truth-telling and the effort tasks, participants had to respond to several test questions to ensure that they understood the tasks and the rules of the experiment. They could continue with the experiment only after those questions were correctly answered.

*1) TRUTH-TELLING TASK.* The truth-telling task consisted of two phases, in each of which the participants were given five choices. Specifically, the task consisted of the following steps: choices in Phase 1, norm manipulation, choices in Phase 2, and a manipulation check.

*Phase 1:* In the truth-telling task, each participant was placed in the situation of a CEO, a manager who had to announce earnings per share for the previous quarter. The participants were told that their salaries had both a fixed and a variable component, the latter being dependent on the announced earnings. They were also told that the market currently anticipated the announcement of 35 cents per share as earnings,<sup>18</sup> but that the true earnings were 31 cents per share.<sup>19</sup> The participants were told that they could announce earnings of 35 cents per share while remaining within legal accounting limits,<sup>20</sup> and that the decision would be solely theirs. They were also informed that they would be paid an amount based on the CEO compensation (that is, on their responses during the decision tasks). This additional experimental payoff was converted into real money at the rate of CHF 100,000 = CHF 0.5.

The participants were then told that they would have to disclose their financial statements that day and that they could choose to announce either 31 or 35 cents as the earnings per share. The two choices were clearly associated with different levels of compensation, reflecting the costs of truth-telling. Specifically, participants were provided with one of two orders of the following tasks:

---

<sup>18</sup> The actual term for the equivalent of cents in the Swiss currency is “Rappen,” and the experiment used the precise Swiss terminology, that is, a choice between 31 Rappen and 35 Rappen, where 100 Rappen = CHF 1. For simplicity we refer to “cents” within the text.

<sup>19</sup> We did not provide the number of shares outstanding. Therefore, there was no explicit statement about the degree of materiality of the earnings per share difference. We thus circumvented the problem, identified by empirical studies, that what practitioners regard as material or immaterial is not generally compatible with what shareholders regard as material or immaterial. In fact, these measures sometimes differ by several orders of magnitude. (See, for example, Seong-Yeon Cho, Robert L. Hagermann, Sandeep Nabar, and Evelyn R. Patterson (2003).) It is, therefore, hard to state presumptively what the right standard would be for our experiment.

<sup>20</sup> 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.

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).

*Norm manipulation:* After Phase 1, we introduced a manipulation based on the notion of Cialdini, Reno, and Kallgren (1990) regarding injunctive norms, that is, beliefs about which behaviors are socially appropriate or inappropriate. For simplicity, we refer to these norms as social norms. The participants were given a page to read that stated that there was the expectation of a good opportunity the following year for the acquisition of another company. However, they would need the shareholders' approval for that project. We then introduced a manipulation by providing some of the participants with additional information. Specifically, the participants were randomly assigned to one of the following three groups:

1. control group

[No further information was provided beyond the common information.]

2. positive social norm group:

"One evening, you are sitting with a friend of yours who is a financial analyst. He tells you that increasing reported earnings in order to meet market expectations meets with widespread approval."

3. negative social norm group:

"One evening, you are sitting with a friend of yours who is a financial analyst. He tells you that increasing reported earnings in order to meet market expectations meets with widespread disapproval."

*Phase 2:* After this interlude, all the participants were again provided with the same set of five options as in Phase 1, requiring them to choose to announce earnings of either 31 or 35 cents per share.

Overall, the participants could earn between CHF 9 and CHF 15 in the two phases of the truth-telling task. It is important to remember that the participants earned less money when they announced the true earnings value of 31 cents. The participants earned the maximum payment when they engaged in earnings management in both choice situations.

*Manipulation check:* Because it was very important to our analysis that the participants correctly identify the choice of the 31-cent option as truthful and of the 35-cent option as deceptive, we included a manipulation check, as is standard in experimental research. For this purpose, the participants were provided with four items and asked, using a 5-point scale ranging from -2 to +2, the extent to which they judged announcing 31 cents as *dishonest vs. honest, manipulative vs. not manipulative, and short-term-oriented vs. long-term-oriented*. The same was also done for the 35-cent option. To verify whether participants correctly perceived opting for 31 cents as personally costly and for 35 cents as personally beneficial, we also asked participants the extent to which they associated announcing 31 cents (or 35 cents) with personal costs vs. personal benefits (-2 = *associated with personal costs* to +2 = *associated with personal benefits*). One final item was designed to assess social concern: We asked participants the extent to which they believed that announcing 31 cents (or 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 potentially relevant control for altruistic preferences or fairness concerns that participants might have and that might confound our inferences regarding protected values of truthfulness.

2) *EFFORT TASK*. Next, participants engaged in a simple calculation task. In this task, each participant was given the role of a manager who could affect the firm's value by working harder. Participants were told that their salaries had both a fixed and a variable component, the latter being dependent on the amount and the accuracy of the work they did. They were then provided with the following slide:

In this task, you can increase earnings per share and, therefore, your compensation, by working. You will work on five sets of calculations. In each set, you can decide whether to do 1 or 5 simple calculations. Doing 5 calculations takes approximately five times as long as doing 1 calculation, and you will be paid more for this. The compensation you receive for 1 and for 5 calculations will vary over the five sets of calculations. Moreover, you will receive CHF 0.2 for each correct calculation.

Participants were shown an example of a calculation, such as  $3 + 4 - 5 + 8 + 3 - 9 = ?$  The participants then read the following screen, one set of choices at a time:

How many calculations do you wish to do?

1 calculation -- In this case, your compensation will be CHF 60,000 (CHF 0.30).

5 calculations -- In this case, your compensation will be CHF 300,000 (CHF 1.50).

1 calculation -- In this case, your compensation will be CHF 120,000 (CHF 0.60).

5 calculations -- In this case, your compensation will be CHF 300,000 (CHF 1.50).

1 calculation -- In this case, your compensation will be CHF 180,000 (CHF 0.90).

5 calculations -- In this case, your compensation will be CHF 300,000 (CHF 1.50).

1 calculation -- In this case, your compensation will be CHF 240,000 (CHF 1.20).

5 calculations -- In this case, your compensation will be CHF 300,000 (CHF 1.50).

1 calculation -- In this case, your compensation will be CHF 300,000 (CHF 1.50).

5 calculations -- In this case, your compensation will be CHF 300,000 (CHF 1.50).

Thus, participants could earn between CHF 4.50 (for always choosing one calculation and doing none correctly) and CHF 12.50 (for doing five correct calculations in each set) on this task.

*3) MEASUREMENT OF PROTECTED VALUES.* The extent to which participants treated truthfulness as a protected value and, therefore, felt committed to truth-telling was then assessed by a measure developed and tested by Tanner, Bettina Ryf, and Martin Hanselmann (2009). In order to ascertain that the answers given by the participants had not been influenced by the experiment, we also asked exactly the same questions in a questionnaire given to individuals who did not participate in the experiment and who did not engage in any tasks other than answering the questionnaire (non-participants).

According to the correspondence (or compatibility) principle established by Icek Ajzen and Martin Fishbein (1980), values and behavior need to be assessed at a similar level of specificity in order to be able to uncover a link between the two. This principle is also followed in the measure developed by Tanner, Ryf and Hanselmann (2009). Since we are studying earnings management behaviors, we adapted the introduction of their questionnaire to the present context. The detailed questionnaire is in the Appendix. The questionnaire contains two subscales, which are highly correlated, but which are designed to approach protected values from different angles. Five items assessed the participants' reactions to hypothetical violations of honesty in reporting company information by a hypothetical CEO. This represents an indirect approach because the decisions of others are being evaluated. The participants and non-participants indicated their judgments about these violations on 7-point

scales, anchored at 1 and 7: (1) *very immoral - very moral*, (2) *not at all praiseworthy - very praiseworthy*, (3) *not at all blameworthy - very blameworthy*, (4) *not at all outrageous - very outrageous*, (5) *not at all acceptable - very acceptable*. These five items allowed us to infer the participants' and non-participants' protected values by examining their moral outrage reactions to norm violations.

Four additional items assessed the participants' own protected values more directly by examining how much importance they attributed to specific features of protected values (such as trade-off reluctance, unwillingness to sacrifice a value, or incommensurability), again referring to the specific context of a hypothetical CEO's decisions regarding the reporting of information. The participants and non-participants were asked to indicate the extent of their personal agreement with the following four statements on a 7-point scale, ranging from *strongly disagree* to *strongly agree*: Truthfulness is about something...(1) that one should not sacrifice, no matter what the (material or other) benefits, (2) for which I think it is right to make a cost-benefit analysis, (3) that cannot be measured in monetary terms, and (4) about which I can be flexible if the situation demands it.

The participants also had to answer another set of questions, which primarily served as control variables. We included questions on their marginal utility of income and on their degree of socially acceptable responding. (See below under control variables for further explanations.) After the experiment, the participants anonymously received their payments of CHF 8 plus their earnings. The average total payment received was slightly less than CHF 30.5.<sup>21</sup>

## **DEPENDENT VARIABLES**

*TRUTHFUL CHOICE*. This represented the dependent variable in the truth-telling 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), while it took on the value of 0 if a participant announced 35 cents (the dishonest option). *TRUTHFUL CHOICE* thus measured the extent of truth-telling, that is, the lack of earnings management.

*EFFORT CHOICE*. This represented the dependent variable in the effort task. It was also coded as a binary variable and took on the value of 1 if a participant chose to do five

---

<sup>21</sup> As explained earlier, by using codes to distribute earnings, we took as much care as possible to ensure anonymity. That is, we tried to remove any possible grounds for expecting reciprocity. It is, therefore, unlikely that a desire to appear honest affected the participants' behavior systematically. See Ariely, Anat Bracha, and Stephan Meier (2009) for a study of how publicly displayed monetary incentives can be less effective in promoting pro-social behavior than privately displayed incentives.

calculations (high effort), while it took on the value of 0 if a participant chose to do one calculation (low effort).

## **EXPLANATORY VARIABLES**

*COST*. This was a within-participant variation. Costs of truthfulness derived from the amount of money a participant forfeited by announcing 31 cents (that is, by truthfulness). As described earlier, participants were twice faced with five situations, from low to high costs associated with truthfulness. The *COST* variable took on values from 1 to 5, with 5 denoting the highest cost of truthfulness (CHF 1.20 = 1.50 – 0.30) and 1 denoting the lowest cost of truthfulness (CHF 0). Thus, an increase by one unit in *COST* corresponded to an additional 30 cents of cost to the participant.

*SOCIAL NORM*. This was a between-participant variation. We defined three dummies, making Phase 1 the omitted category in the regressions. *CONTROL* was equal to 1 for all observations from Phase 2 with no additional information, and to 0 otherwise. *POSITIVE NORM* was equal to 1 for all observations from Phase 2 with the positive social norm (acceptance of earnings management), and to 0 otherwise. *NEGATIVE NORM* was equal to 1 for all observations from Phase 2 with the negative social norm (disapproval of earnings management), and to 0 otherwise.

*PROTECTED VALUES (PV)*. After appropriate recoding of some items, an index of the degree of protected values (*PV*) was constructed, based on the means across all nine direct and indirect items. This index measured the extent to which the participants treated truthfulness as a protected value and felt therefore committed to truth-telling; it took on a value between 1 (for an individual with no protected values) and 7 (for an individual with maximum protected values). The internal consistency of this scale, as assessed by Cronbach's  $\alpha$ , was very satisfactory ( $\alpha = 0.86$ ).<sup>22</sup>

*ALTRUISTIC CONCERNS*. The extent to which participants judged that announcing 35 cents would hurt other stakeholders was coded as the variable *35HURTS*. This item measured the

---

<sup>22</sup> Cronbach's Alpha is an important 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. We also did the analysis using the direct and indirect sub-scales separately. Both sets of results were quantitatively similar and are available upon request.

extent to which the participants believed that their earnings management had negative consequences for other stakeholders.

## CONTROL VARIABLES

The participants had to answer several questions in order for us to assess some of their individual characteristics. These included questions on the marginal utility of income, on socially acceptable responding, and on demographics.

*MARGINAL UTILITY OF INCOME.* We asked the following question:<sup>23</sup>

Please imagine that you find a CHF 50 bill on the street. It is impossible to identify the owner, and it is, therefore, completely acceptable and morally unobjectionable that you keep the CHF 50. Think about your average peer who earns about the same amount of money as you do, and is approximately equally wealthy. Would you say that, relative to this average peer, you benefit

- \_\_\_ a lot more
- \_\_\_ more
- \_\_\_ equally
- \_\_\_ less
- \_\_\_ a lot less

from this additional amount of money?

We assigned a value of 5 to “a lot more” answers, and a value of 1 to “a lot less” answers. This measure captures each participant’s self-reported marginal utility of income.

*SOCIALLY ACCEPTABLE RESPONDING.* We used the standard Deception Scales (PDS) of Delroy Paulhus (1984); see Jochen Musch, Robbi Brockhaus, and Arndt Bröder (2002) for the German version. This is a self-reporting questionnaire designed to measure individuals’ tendencies to give socially desirable responses (SDR). It measures two distinct forms of SDR: self-deception and impression management. Accordingly, we coded two variables *SELFDECEIT* and *EXTDECEIT*. Participants who exhibited more socially acceptable responses scored higher on both scales.

*DEMOGRAPHIC CONTROL VARIABLES.* *SEX*, *AGE*, and *STUDIES* served as 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

---

<sup>23</sup> In some societies, it is unacceptable to keep found money. For example, in Japan one is expected to bring it to a police station. We had no students from such societies in our sample.

participants from fields other than psychology and economics and to 0 otherwise. *ECONOMICS* was the omitted category.

## **SEPARATE SURVEY WITH NON-PARTICIPANTS**

We also conducted a survey with freshmen university students who had not participated in any of the previously described experiments. We only measured the protected values of these students and did not involve them in any of the choice tasks. Thus, we were able to compare the distributions of protected values in the experimental sample and in a sample of non-participants who could not have been influenced by any elements of the actual experiment.

## **4. Experimental results**

We began by verifying that our participants perceived the announcement of 31 cents as the honest, non-manipulative action that led to a personal loss, while the opposite was true of the announcement of 35 cents. As can be seen in Table I, the results of the manipulation check strongly confirm this expectation.

---

TABLE I ABOUT HERE

---

### ***A. Sample characteristics***

Table II provides descriptive statistics and, together with Table III, allows a first look at the choices the participants made in the experiment.

Having pooled all situations and manipulations, we found that, in approximately 42% of cases, participants chose to announce low earnings, that is, chose not to engage in earnings management, opting instead to suffer monetary losses relative to what they could have earned. Table III also reveals that there was substantial variation in responses throughout the within-participants and between-participants conditions that were established in the experiment.

The fact that a large proportion of the participants reported the truth, even when the conditions opposed it, is consistent with the findings of the earlier studies mentioned in the Introduction. It is important to recall that, 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 (we had told the participants explicitly that earnings management was legal), and no risk of being punished.

Even when there was no cost for truthfulness, 23% of the participants chose the earnings management solution. It appears difficult to rationalize their behavior with either the opportunistic or the protected values model. The manipulation check also confirmed that this group of people saw 35 cents as the less honest option (though the difference was indeed smaller than in the rest of the sample).

---

TABLE II ABOUT HERE

---

---

TABLE III ABOUT HERE

---

This is the point at which much of the existing literature stopped. We used, as the starting point for our inquiry, the observation that a large proportion of people reported the truth even when it was to their disadvantage. Exploring this phenomenon further was the main thrust of this study.

### ***B. Protected values***

As Table II reveals, we found substantial variation in the extent to which participants stated their commitments to the protected value of truthfulness. The average of *PV* was 4.82 (out of 7), while the 1%, 10%, 25%, 75%, 90%, and 99% percentiles were 2.6, 3.6, 4.0, 5.6, 6.1, and 7.0, respectively.

As mentioned in the previous section, we also conducted the same survey with 123 students from another class who did not participate in this experiment. The distribution of protected values in this group looked virtually identical to the one we obtained in the experimental group. (See the last row of Table II.) This increased our confidence in the measure of the protected value of truthfulness, and it also made it highly unlikely that the answers our participants gave to the protected values questions were affected by the experimental setting.

### ***C. Truthful choice and values: Univariate evidence***

We next considered the univariate relationship between protected values of truthfulness and the earnings announcement decisions made by participants. Figure 1 shows two density plots clarifying this relationship.

In this Figure, we plot the densities of the protected values of truthfulness separately for those participants who chose 31 cents and for those who chose 35 cents. In order to isolate the role of protected values, the graph plots only data from Phase 1 of the experiment and only from the median cost level. The message from this picture is clear: Those who chose to forgo monetary rewards in favor of reporting the truth were much more likely to be participants with strong protected values. By contrast, participants who managed earnings in order to obtain higher pay were much more likely to possess medium or weak protected values. This graph looks very similar to the plotted data at the other cost levels. The average *PV* of those who never lied was 5.78 (out of 7), as opposed to 4.82 in the overall sample, and 4.38 for those who always lied.

#### ***D. Truthful choice and values: Controlling for other factors***

To study the role of monetary costs and of social norms on agents' choices, we estimated (subsets of) the following basic model:

$$\Pr(\text{TRUTHFULCHOICE}_{ict}) = F \left[ \begin{array}{l} a + b * PV_i + c * \text{COST}_{ct} + d * \text{CONTROL}_{it} + \\ + e * \text{POSITIVE NORM}_{it} + f * \text{NEGATIVE NORM}_{it} + \\ + g * \text{DEMOGRAPHICS}_i \end{array} \right] \quad (1)$$

where the index  $i$  refers to the participant, where  $c=1, \dots, 5$  refers to the explicit economic cost of truthfulness (as measured by the income forgone by the participant when he or she announced 31 cents of earnings), and where  $t=1$  or  $2$  and refers to the phase of the experiment. Note that the Phase 1 dummy was omitted, so the coefficients  $d$ ,  $e$ , and  $f$  were interpreted relative to the Phase 1 choice. We considered the full set of participants,  $i=1, \dots, 261$ , but we also separately studied participants divided into terciles (and higher quantiles) of *PV*, with correspondingly smaller sample sizes. For the main analysis, we used logit estimates, in which case  $F[\cdot]$  is the standard normal cumulative distribution function. (Using probit yielded similar results.) Recall that in each phase, participants went through all five cost situations, creating ten observations for each participant. The regressions correct the standard errors for possible serial correlation and heteroskedasticity by clustering at the individual level.

Table IV presents the regression results for several specifications. Column 1 confirms the earlier finding, showing that the participants with stronger protected values of truthfulness were much less inclined to engage in earnings management. The following columns add other

explanatory variables that the theory suggests, as well as control variables. A number of results are noteworthy.

---

TABLE IV ABOUT HERE

---

First, no matter which control variables we included, protected values for truthfulness remained an important, statistically significant determinant of earnings management decisions. Calculating marginal effects from these regressions, we find that a one-point increase in *PV* was associated with an expected increase in the probability of no earnings management of about 20 percentage points, holding the other variables at their means.

Second, the monetary cost of truthfulness mattered, and it did so in the way expected from the preference specification presented in Section 2. Table III has already suggested this, and Table IV now formally tests for this, first focusing only on Phase 1 (Column 3), then on Phase 2 (Column 4), and then combining the two phases for an analysis of the joint impact of *COST* and norms (Column 5). In all three regressions, we found that the more a participant would lose when announcing low earnings, the more the participant was likely to announce high earnings, as can be seen from the negative coefficients of *COST*. Evaluating the marginal effects at the mean, a one-cent increase in the cost was associated with somewhat more than a half-percentage-point decrease in the probability of truthfulness. Note that this effect persisted after controlling for protected values. In other words, in the cross-section of participants, the presence of a strong value of truthfulness did not completely eliminate the relevance of the situation in which the earnings management decision was made. Note also that while Hurkens and Kartik (2009) demonstrated that Gneezy's (2005) data would be consistent with a population of pure opportunists and pure ethical types where the former always lie and the latter always tell the truth (as in Koford and Penno 1992), the present evidence is directly in contrast to the implications of such a types-based model. Another way to understand this finding is to note that there were 7 participants with supposedly "absolute" protected values,  $PV = 7$ . Of these, 3 never engaged in earnings management in any of the 10 situations. Notably, there were 42 other participants who also always told the truth. The protected values concept acted as a useful organizing principle; but at least with respect to truth-telling, an absolute stated commitment to the value of truthfulness was neither a necessary nor a sufficient condition for avoiding earnings management.

The third finding—already suggested in Table III—is that participants did react to the information we provided in Phase 2 of the experiment, namely, whether society either approved

of or disapproved of earnings management. Expressing the coefficients on the *POSITIVE NORM*, *NEGATIVE NORM*, and *CONTROL* dummies, respectively, as marginal effects implies that the positive social norm made earnings management 15% more likely, the negative social norm made it 15% less likely, and the control group behaved, on average, about the same as in Phase 1. These results are in line with the predictions on the general impact of social norms outlined in Section 2.

We found no robust additional effect from the ages, genders, or fields of study of the participants in our experiment.

### ***E. Interactions of values with costs and social norms***

So far we had established that many participants reported the truth in a situation where most economists would not have expected it and that the extent of truth-telling by decision makers was correlated with the value of truthfulness to these persons. We also wanted to test finer implications of the model of choice which we described in Section 2. In particular, we first evaluated the proposition that, even if the costs mattered to all participants, varying them would have less impact on the choices made by the participants with strong protected values of truthfulness.

We now structure our discussion along three pieces of evidence. First, Figure 2 considers the levels of truthfulness of participants with different *PVs* in the five different Phase 1 *COST* situations. Second, Figure 3 Panel A highlights the differing levels of steadfastness of the participants in the face of *COST* changes. Third, regressions allow us to consider the statistical significance of the differences in resistance to economic costs.

Figure 2 provides a look at behavior in Phase 1, in order to isolate the interaction of protected values with monetary costs and to abstract away from the influence of norms. Figure 2 categorizes participants along two dimensions: the *PV* tercile to which they belong and their behavior in the experiment. Within each category of *PV*, the percentages in the various behavior categories add up to 100%. Figure 2 shows that those in the lowest tercile of *PV* were largely telling the truth only when it was free, and fewer than 5% of them always told the truth. By contrast, relatively few (around 20%) of those in the top tercile of *PV* could be found in the group of those who told the truth only when it entailed no cost, and almost 40% of them always told the truth.

---

FIGURE 2 ABOUT HERE

---

Panel A of Figure 3 offers another condensed perspective on the same data. This graph demonstrates that the higher a participant's *PV*, the less likely it was that an increase in direct truth-telling costs induced that participant to switch from truthfulness to lying, that is, the more steadfast truthfulness that participant demonstrated. Averaging over all cost situations, for those in the lowest tercile of *PV*, only around 50% of those who told the truth at a given cost level continued to keep telling the truth at the next highest cost level. By contrast, measured in this way, around 80% of those in the highest tercile of *PV* were resistant to the detrimental influence of the economic costs of truthfulness.

---

FIGURE 3 ABOUT HERE

---

These results are also reflected in our logit regressions, as presented in Table V. In Column 1 of this table, we obtain a positive coefficient on the interaction term between *PV* and *COST*. While the interpretation of interaction terms in regressions with limited dependent variables is difficult, this provides suggestive evidence that, for participants with strong protected values, an increase in *COST* was associated with a smaller increase in the probability of lying than for those with weaker values. To make further progress, we adopted a non-parametric approach and next considered the main regression separately for the three terciles of *PV*. These results are presented in Table V in Columns 2 - 4. At the bottom of the table, we present differences of coefficients between the highest and lowest *PV* quantile under consideration and the associated z-statistics for the significance of these differences.

Note that to test the theoretical predictions regarding the differential sensitivities to economic costs of participants with different *PVs*, we consider coefficients, rather than marginal effects, from the logit regressions. It would not be surprising to find a small marginal effect of *COST* for those with strong protected values: They were initially (before a change of economic costs) more likely to report the truth. Analyzing coefficients instead allows us to consider the hypothetical case of participants who would display identical initial probabilities of reporting the truth.

---

TABLE V ABOUT HERE

---

The central result is that the participants whose *PV* placed them in the top tercile of *PVs* proved much less influenced by the cost increase than those in the lowest tercile. The

difference in coefficients between the two terciles is highly statistically significant.<sup>24</sup> Columns 5 and 6 further demonstrate that, for the highest quantile in other sample splits, the sensitivity to *COST* was also very small (and decreased in the higher *PV* quantiles). With such fine-grained sample splits, the difference between two adjacent *PV* quantiles was not generally significant, but participants with weaker-than-median and stronger-than-median protected values reacted significantly differently (not shown).

We also confirmed (in untabulated results) that, indeed, the highest marginal effects of an increase in the costs of truth-telling occurred in the middle *PV* group. Underlying the small marginal effects for the bottom tercile is the fact that those with low *PVs* had a high initial probability of lying, and so a change in the costs of truthfulness did not alter their behavior. Those with strong *PVs* were so far above the truth-telling / lying threshold that changes in economic costs had very little effect on their behavior.<sup>25</sup>

In sum, the body of evidence presented so far strongly supports the predictions of the preference specification as described in Section 2. The result, that those with higher protected values were more likely to report the truth, is consistent with the idea (captured by  $c(PV_i) > 0$  in the preference specification) that participants had a preference for truth per se. The finding that those with strong protected values were particularly impervious to the economic costs of truthfulness is consistent with the notion (captured by the discount parameter  $k(PV_i) > 0$  in the preference specification) that the participants' perception of funds received depended on whether these receipts were due to reporting the truth or to lying, and that those with stronger values discounted funds received by lying more.

We considered next the differential impact of social norms among participants with different levels of protected values for truthfulness. Again, we structure our discussion around three pieces of evidence, one graph highlighting the different changes in truthfulness by different participants, a second graph showing the levels of truthfulness, and a set of regressions. Specifically, first, in Figure 3, Panels B and C show any changes in behavior from

---

<sup>24</sup> A perhaps interesting effect is that in the median *PV* range, women were more likely to tell the truth. The fractions of men and women in the median and top third *PV* groups were very similar. Another point to note about Table V is that the results also hold when including *PV* in the regressions for the separate quantiles, though *PV*'s explanatory power naturally declines as we increase the number of quantiles considered. When we do include *PV*, we can strongly reject, with a likelihood ratio test, the hypothesis that the coefficients in models 2 – 4 in Table V are identical to those associated with model 5 in Table IV.

<sup>25</sup> Formally, recall that marginal effects in a logit regression are given by  $\Lambda(\beta' \mathbf{X})(1 - \Lambda(\beta' \mathbf{X}))\beta$ , where  $\Lambda(\bullet)$  is the logistic cumulative distribution function giving the initial probability of truthfulness, and  $\beta$  is the vector of coefficients. Those with strong protected values have high initial probabilities of truthfulness, those with weak values have low initial probabilities. Thus, the highest marginal effects are likely to be found in the middle range.

Phase 1 to Phase 2, thus using a participant's earlier behavior as a reference point. Second, Figure 4 plots the levels of percentages of truth-tellers in Phase 2 across the three norms treatments (over all *COST* situations), now using the *CONTROL* group as reference point. Recall that three norms treatments were allocated randomly. Third, the regressions in Table V allow us to consider both phases jointly and to control for both *COST* and social norms.

---

FIGURE 4 ABOUT HERE

---

The primary prediction that flows from our preference specification, as laid out in Section 2, is that those with strong *PVs* will react less to norms. In line with this prediction, Figure 4 suggests that—relative to the percentage of truth-tellers in the *CONTROL* condition—for both the *POSITIVE* and the *NEGATIVE NORMS* there was much less variation in behavior among the top third *PV* group than among the bottom third *PV* group. Panel B of Figure 3 documents that the imperviousness to the prodding of the *POSITIVE NORM* (approval of earnings management) was greater among those with medium or strong *PVs* than among those with weak *PVs*. For the *NEGATIVE NORM* treatment, by contrast, Panel C of Figure 3 suggests that in the within-participant comparison, the steadfastness of participants was similar in the three terciles.

This somewhat asymmetric picture with respect to the positive and negative norms is also reflected in the regressions. The statistically significant interaction term *PV \* POSITIVE NORM* in Column 1 of Table V suggests that strong protected values countered some of the increased lying effect that would otherwise have been due to a positive social norm. The statistically insignificant interaction term *PV \* NEGATIVE NORM* instead suggests that those with strong protected values were not so unaffected by a disapproving social norm.<sup>26</sup> Comparing coefficients in columns 2 to 6 confirms these observations: Participants with strong protected values were generally more impervious to the positive norm. For the negative norm, this higher imperviousness only becomes statistically significant at the more extreme quantiles.<sup>27</sup>

---

<sup>26</sup> Comparing coefficients on *COST* in regressions investigating separately the *POSITIVE NORM* and *NEGATIVE NORM* conditions with those in the *CONTROL* condition, we did not find evidence that either type of social norm affected the participants' responses to the economic costs of truthfulness. This is consistent with the main specification of the global utility function in Section 2 which postulates that social norms tend to work primarily through the separate term capturing the moral and social costs of lying,  $c_i$ .

<sup>27</sup> A caveat to the overall results regarding social norms is that those in the lowest quantile tended to lie more in the *CONTROL* condition than in Phase 1, while those in the highest quantile tended to report the truth more. The difference is not statistically significant at conventional levels, but as can be seen in Column 5 of Table V, in the case of the quartile split (but not in other splits), it is about as significant as the difference in reactions to the *POSITIVE* and *NEGATIVE NORMS*. Thus, it is conceivable that part of the strong reaction to the *POSITIVE*

The regressions also imply that the marginal effect on the probability of truth-telling of positive and negative social norms was greatest for participants whose protected values approximated the median (not shown).<sup>28</sup>

To summarize the findings from this second phase of the experiment, in accordance with the preference specification in Section 2, participants with strong protected values generally reacted less to social norms, especially when the norm was supportive of earnings management. While the powerful effects of social norms on average behavior are known in the psychology literature, our study shows that there exist some individuals who can, in fact, resist such norms, and that this resistance derives from their protected values of truthfulness. These results are, however, not quite as statistically significant as those regarding their steadfastness in the face of the economic costs of truthfulness.

## ***F. The effort task***

We also wanted to examine whether the role played by protected values in the choice of whether or not to manage earnings could have been at least partially driven by experimenter demand. That is, one potential concern was that participants might somehow have inferred that the experimenters associated their choices of the truthful option with their protected values of truthfulness, causing their responses on both the survey and on the truth-telling task to have been in line with what they thought might be the experimenters' expectations. In order to guard against this conjecture, we also engaged the participants in performing a simple effort task in which they exerted more or less effort, depending on the number of simple arithmetic computations that they chose to perform. To the extent that the participants were also responding to experimenter demand in this task, those who had expressed stronger protected values would also do more calculations. By contrast, under the null hypothesis, protected values would be causally linked only to truth-related choices, and *EFFORT CHOICE* would not in any way be related to a participant's protected value of truthfulness.

The evidence presented in Figure 5 serves as an indication that the previously established positive relationship between the participants' degrees of protected values and their truth-telling choices was not driven by an experimenter-demand effect. In particular, this figure

---

*NORM* of those with weak values was caused by a general effect decreasing those participants' truthfulness in Phase 2. If so, though, the results would also imply that imperviousness to the *NEGATIVE NORM* was, in fact, significantly higher among those with strong protected values.

<sup>28</sup> Intuitively, for those who were strongly opportunistically inclined, social norms regarding earnings management did not have measurable behavioral effects, either because a positive norm encountered people who were already lying or because a negative norm failed to dislodge participants who were initially too unmotivated to consider truthfulness as a viable option.

shows that there was no observable relationship between the participants' levels of effort in performing this task and their levels of protected values.<sup>29</sup>

---

FIGURE 5 ABOUT HERE

---

### ***G. Further analysis and robustness***

We further explored the roles of (1) the marginal utility of income, (2) altruistic concerns, (3) socially acceptable responding and external deceit, (4) stock ownership, (5) perceptions of CEOs in newspaper articles, (6) alternative quantiles, and (7) the use of samples with strictly positive costs of truthfulness only, alternative estimation methods, and randomizations in the experiment. Table VI shows the results for features 1, 2, 3, and 6 (by focusing on the four quartiles of *PV*), while we comment only briefly on the other issues. To conserve space, Table VI only presents the coefficients (and differences of coefficients for the top and bottom *PV* quartiles) of the key variables of interest, but all regressions include the usual controls.

---

TABLE VI ABOUT HERE

---

*Marginal utility of income.* One potential concern is that the participants might simply have reported the truth because it did not cost them much relative to their income or wealth, or because they did not care about money. On a general level, it is well-accepted that small stakes are informative in experiments. Nonetheless, we wished to test this idea in a specific way. Therefore, we included the measure of marginal utility that we defined for each participant in the regressions. In the equivalent of Regression 5 of Table IV, this measure was insignificant (not shown). For the bottom quartile *PV* group in Table VI, Panel A, the coefficient was significant and negative, indicating that among those with weak protected values, richer participants (or those caring less about money) reported the truth more often. But importantly, the difference in resistance to *COST* and social norms remained at a similar level as before when we controlled for marginal utility. Also, the correlation between marginal utility and protected values was virtually zero. It is important to recall that the protected values measure does not imply that participants do not care about money; it only measures how much less willing they are to make trade-offs.

---

<sup>29</sup> This finding is further confirmed with high statistical significance in a regression analysis similar to the one performed for the truth-telling task. To avoid cluttering Figure 5, we present the results for above- and below-median *PVs* only. The figures for three, four, or more *PV* groups look very similar.

*Altruistic concerns.* The extent to which participants perceived the 35-cent option to be damaging to stakeholders did not vary between participants from the fields of economics and of psychology. Women perceived the 35-cent option as less harmful to other stakeholders than men did, but women still saw it as significantly more harmful than the 31-cent option. *35HURTS* has a positive correlation with *PV* of 0.34, but including the altruistic concerns variables *35HURTS* in the regressions yielded a significant effect for some *PV* quantiles, suggesting that the two concepts are different. (See, for example, Column 3 of Table VI, Panel B). The previous main results continued to hold, with slightly higher significance for the differential resistance to social norms. Finally, it should be noted that, on a conceptual level, *35HURTS* should not play a role in this experiment since it was designed to rule out social interactions. By contrast, the interpretation of the results for *PV* is consistent in light of the experimental setup chosen for this study.

*Socially acceptable responding.* When we control for socially acceptable responding, the difference in reactions remains as before for *COST*, becomes somewhat less significant for the *POSITIVE NORM*, and is again insignificant at conventional levels for the *NEGATIVE NORM*. Given the design of the experiment, in which we took great care to make the responses anonymous, it is very unlikely that participants developed an interest in impressing the experimenter by appearing honest. Participants received a code with which they were able to pick up the funds they had earned, and the earnings were pooled and allocated for two distinct tasks, the truth-telling and the effort tasks. Moreover, if anything, we would have expected those with a high preference for socially acceptable behavior to have reacted more strongly to the social norms we presented. However, we tended to observe exactly the opposite: Those with strong protected values for truthfulness generally reacted less to social norms.

*Validity outside of the laboratory.* We asked the participants whether they had recently read newspaper articles about CEOs, and, if so, whether the CEOs had been depicted in a positive or negative light. Neither a participant's exposure to articles about CEOs nor the quality of the CEOs' depictions had statistically significant influence on the choices made in the values-based task in the experiment. We also asked the participants whether they owned stocks (as individual holdings or through mutual funds). This variable was not correlated with any of the choice variables and was not significant in the regressions.

*Sample and estimation methods.* We considered many other variations of the analysis. These include: (a) omitting the participants aged 34 and older (that is, the 95<sup>th</sup> percentile and above in age), which did not change the results; (b) omitting the situation in which the truth was cost-free, which did not materially affect the results; (c) comparing the behaviors of the

participants in the cost randomization and the timing randomization, in which we found no statistically significant differences; and (d) using, as already mentioned, random effects logit regressions as an alternative estimation procedure, with results very similar to those from before, except for one difference concerning gender. Using this estimation method, the gender dummy was significantly positive in most regressions with other controls, and it was always significant when we controlled for the monetary costs of truthfulness (independent of whether we also controlled for social norms). Thus, women seemed to separate from their male counterparts when we controlled for the extent to which a truthful choice required a sacrifice. Since this effect was only obtained when using one of the two estimation methods, this finding cannot be seen as robust; however, it merits further inquiry.

## **5. Implications and conclusion**

In this study, we established that some individuals have a strong, non-selfish, non-strategic, and non-consequentialist commitment to truthfulness, and that there is substantial heterogeneity among individuals with respect to this commitment. We used the concept of protected values to refer to the well-established psychological foundations of this commitment. We hypothesized that the protected value of truthfulness can powerfully affect economic decision making, and we tested this hypothesis in an experiment. We used the fairly common practice of earnings management in the financial world to cast the experiment in a concrete business setting.

The main findings are the following. First, in 42% of the situations, participants chose to forgo additional compensation in favor of reporting the truth about the earnings of their firms. This is a surprisingly high number, at least judging by the utility-maximizing paradigm. However, this finding is consistent with our hypothesis of agents' choices being driven by a heterogeneous commitment to truthfulness.

Additionally, the protected value of truthfulness varied among our participants, was rarely absolute, and, importantly, had strong explanatory power for a wide range of behaviors on the honesty spectrum. This finding complements and illuminates existing research showing that individuals are frequently motivated by factors other than pure self-interest.

Further, the power of most of our participants' commitments to truthfulness was in general negatively influenced by the compensation that the participants had to forgo if they decided to report the truth. However, those participants with stronger protected values for truthfulness

were generally less responsive to changes in the monetary benefits of lying and less influenced by social norms.

Finally, we demonstrated that the measured protected values of truthfulness had no predictive power in a pure effort task unrelated to values, thus supporting a causal relationship between the agents' commitments to truthfulness and the economic choices they made in the truth-telling task. Further, a large set of robustness checks was performed, bolstering confidence in these experimental results and in their external validity. For example, the distribution of protected values in a sample of non-participants was very similar to the distribution in our sample of participants. Additionally, protected values were not correlated with the stated marginal utility of income.

The evidence thus leads us to conclude that we cannot reject the hypothesis that a protected-value-based motive for truth-telling exists, that it can be measured, that it varies among individuals, that its impact varies across monetary and social situations, and that it meaningfully influences economic decision-making. Some existing studies have inferred, from apparently “excessive” truth-telling behavior that cannot be explained with standard economic models, that there must exist preferences for truthfulness. However, those studies did not measure those preferences at the individual level, nor did they test the finer predictions of such a hypothesis in an experimental setting that rules out social interactions. By exploring these avenues, this study provides direct evidence of a heterogeneous commitment to the protected value of truthfulness and of its influence on economic decisions.

Our findings bear relevance for the current discussions among regulators, investors, managers, and board members of financial and industrial firms regarding the reshaping of managerial remuneration policies and the selection of agents.<sup>30</sup> There is ample room for further research to spell out the actual workings, implications, and limitations of a system of governance that builds on protected values as one of the driving forces behind economic decision making.

---

<sup>30</sup> For example, our observation that agents with a strong commitment to truthfulness were less sensitive to, and thus less motivated by, financial incentives is consistent with the recent theoretical argument by Carlin and Gervais (2008), which led to their proposing the introduction of “ethical fixed-wage contracts” to help principals hire virtuous agents. A labor market view, as in Timothy Besley and Maitreesh Ghatak (2005), can shed some light on how organizations aim to attract those employees whose values fit with one another. See also Michael Kosfeld and Ferdinand von Siemens (2007) for an analysis of endogenous sorting of workers by their cooperative attitudes. Finally, protected values are likely to affect the dynamic relationship between business activity and ethics, as set forth in Noe and Rebello (1994).

## 6. References

- Ajzen, Icek, and Martin Fishbein.** *Understanding Attitudes and Predicting Social Behavior*. Englewood Cliffs: Prentice-Hall, 1980.
- Ariely, Dan, Anat Bracha, and Stephan Meier.** 2009. "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially" *American Economic Review*, forthcoming.
- Atran, Scott, Robert Axelrod, and Richard Davis.** 2007. "Sacred Barriers to Conflict Resolution." *Science*, 317: 1039-40.
- Baiman, Stanley, and Barry L. Lewis.** 1989. "An Experiment Testing the Behavioral Equivalence of Strategically Equivalent Employment Contracts." *Journal of Accounting Research*, 28(1): 1-20.
- Baron, Jonathan, and Sarah Leshner.** 2000. "How Serious Are Protected Values?" *Journal of Experimental Psychology: Applied*, 3: 183-94.
- Baron, Jonathan, and Mark Spranca.** 1997. "Protected Values." *Organizational Behavior and Human Decision Processes*, 70(1): 1-16.
- Benabou, Roland, and Jean Tirole.** 2010. "Identity, Dignity and Taboos: Beliefs as Assets." *Quarterly Journal of Economics*, forthcoming.
- Bergstresser, Daniel, and Thomas Philippon.** 2006. "CEO Incentives and Earnings Management." *Journal of Financial Economics*, 80(3): 511-29.
- Besley, Timothy, and Maitreesh Ghatak.** 2005. "Competition and Incentives with Motivated Agents." *American Economic Review*, 95(3): 616-36.
- Bhide, Amar, and Howard H. Stevenson.** 1990. "Why Be Honest If Honesty Doesn't Pay." *Harvard Business Review*, 68(5): 121-29.
- Carlin, Bruce, and Simon Gervais.** 2008. "Work Ethic, Employment Contracts, and Firm Value." *Journal of Finance*, forthcoming.
- Cho, Seong-Yeon, Robert L. Hagermann, Sandeep Nabar, and Evelyn R. Patterson.** 2003. "Measuring Stockholder Materiality." *Accounting Horizons*: 63-76.
- Cialdini, Robert B., Carl A. Kallgren, and Raymond R. Reno.** 1991. "A Focus Theory of Normative Conduct." *Advances in Experimental Social Psychology*, 24: 201-34.
- Cialdini, Robert B., Raymond R. Reno, and Carl A. Kallgren.** 1990. "A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places." *Journal of Personality and Social Psychology*, 58: 1015-26.
- Davis, Douglas D., and Charles A. Holt.** *Experimental Economics*. Princeton: Princeton University Press, 1992.
- DeJong, Douglas V., Robert Forsythe, and Wilfred C. Uecker.** 1988. "A Note on the Use of Businessmen as Subjects in Sealed Offer Markets." *Journal of Economic Behavior and Organization*, 9: 87-100.
- Dyer, Douglas, John H. Kagel, and Dan Levin.** 1989. "A Comparison of Naive and Experienced Bidders in Common Value Offer Auctions: A Laboratory Analysis." *Economic Journal*, 99: 108-15.
- Ernst & Young.** "European Fraud Survey 2009: Is Integrity a Casualty of the Downturn?," 2009.
- Evans, John H., R. Lynn Hannan, Ranjani Krishnan, and Donald V. Moser.** 2001. "Honesty in Managerial Reporting." *Accounting Review*, 76(4): 537-59.
- Fehr, Ernst, and Urs Fischbacher.** 2002. "Why Social Preferences Matter." *Economic Journal*, 112(478): C1-C33.
- \_\_\_\_\_. 2003. "The Nature of Human Altruism." *Nature*, 425(23 October 2003): 785-91.

- Fehr, Ernst, and John A. List.** 2004. "The Hidden Costs and Returns of Incentives - Trust and Trustworthiness among CEOs." *Journal of the European Economic Association*, 2(5): 743-71.
- Fischbacher, Urs, and Franziska Heusi.** 2008. "Lies in Disguise: An Experimental Study on Cheating." *Mimeo*.
- Fischer, Paul, and Steven Huddart.** 2008. "Optimal Contracting with Endogenous Social Norms." *American Economic Review*, 98(4): 1459-75.
- Forsythe, Robert, Joel L. Horowitz, N. E. Savin, and Martin Sefton.** 1991. "Fairness in Simple Bargaining Experiments." *Games and Economic Behavior*, 6: 347-69.
- Gillette, Ann B., Thomas H. Noe, and Michael J. Rebell.** 2008. "Board Structures around the World: An Experimental Investigation." *Review of Finance*, 12: 93-140.
- Gneezy, Uri.** 2005. "Deception: The Role of Consequences." *American Economic Review*, 95(1): 384-94.
- Graham, John R., Campbell R. Harvey, and Shiva Rajgopal.** 2005. "The Economic Implications of Corporate Financial Reporting." *Journal of Accounting and Economics*, 40: 3-73.
- 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(October): 365-83.
- Hoffman, Elizabeth, Kevin A. McCabe, and Vernon Smith.** 1996. "On Expectations and the Monetary Stakes in Ultimatum Games." *International Journal of Game Theory*, 25(3): 289-301.
- Hurkens, Sjaak, and Navin Kartik.** 2009. "Would I Lie to You? On Social Preferences and Lying Aversion." *Experimental Economics*, 12(2): 180-92.
- Jensen, Michael C.** 2007. "Putting Integrity into Finance Theory and Practice: A Positive Approach." *Mimeo*.
- Koford, Kenneth, and Mark Penno.** "Accounting, Principal-Agent Theory, and Self-Interested Behavior," N. E. Bowie and R. E. Freeman, *Ethics and Agency Theory: An Introduction*. New York: Oxford University Press, 1992.
- Kosfeld, Michael, and Ferdinand von Siemens.** "Competition, Cooperation, and Corporate Culture," 2007.
- Mazar, Nina, On Amir, and Dan Ariely.** 2008. "The Dishonesty of Honest People: A Theory of Self-Concept Maintenance." *Journal of Marketing Research*, XLV: 633-44.
- Mittendorf, Brian.** 2007. "Infectious Ethics: How Upright Employees Can Ease Concerns of Tacit Collusion." *Journal of Law, Economics and Organization*, 24(2): 356-70.
- 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*, (Supplement): 17-35.
- Noe, Thomas H., and Michael J. Rebell.** 1994. "The Dynamics of Business Ethics and Economic Activity." *American Economic Review*, 84(3): 531-47.
- Paulhus, Delroy.** 1984. "Two-Component Models of Socially Desirable Responding." *Journal of Personality and Social Psychology*, 46: 598-609.
- Rabin, Matthew.** 1993. "Incorporating Fairness into Game Theory and Economics." *American Economic Review*, 83(5): 1281-302.
- \_\_\_\_\_. "Moral Preferences, Moral Constraints, and Self-Serving Biases," 1995.
- Rode, Julian.** 2010. "Truth and Trust in Communication - Experiments on the Effect of a Competitive Context." *Games and Economic Behavior*, 68: 325-38.

- Sade, Orly, Charles Schnitzlein, and Jaime F. Zender.** 2006. "Competition and Cooperation in Divisible Good Auctions: An Experimental Investigation." *Review of Financial Studies*, 19: 195-235.
- Sánchez-Pagés, Santiago, and Marc Vorsatz.** 2007. "An Experimental Study of Truth-Telling in a Sender-Receiver Game." *Games and Economic Behavior*, 61: 86-112.
- Shleifer, Andrei.** 2004. "Does Competition Destroy Ethical Behavior?" *American Economic Review*, 94: 414-18.
- Sims, Randi L.** 2000. "The Relationship between Employee Attitudes and Conflicting Expectations for Lying Behavior." *Journal of Psychology*, 134(6): 619-33.
- Smith, Vernon, Gerry L. Suchanek, and Arlington W. Williams.** 1988. "Bubbles, Crashes, and Endogenous Expectations in Experimental Spot Asset Markets." *Econometrica*, 56: 1119-51.
- Sutter, Matthias.** 2009. "Deception through Telling the Truth?! Experimental Evidence from Individuals and Teams." *Economic Journal*, 119: 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: 757-69.
- Tanner, Carmen, Bettina Ryf, and Martin Hanselmann.** 2009. "Geschützte Werte Skala: Konstruktion und erste Validierung eines Messinstrumentes (Sacred Value Measure: Construction and First Validation of an Instrument to Assess Sacred Values)." *Diagnostica*, 55(3): 174-83.
- Terry, Deborah J., and Michael A. Hogg.** "Attitudes, Behavior, and Social Context: The Role of Norms and Group Membership in Social Influence Processes," Joseph P. Forgas and Kipling D. Williams, *Social Influence: Direct and Indirect Processes*. Psychology Press, 2001.
- Tetlock, Philip E.** 2003. "Thinking the Unthinkable: Sacred Values and Taboo Cognitions." *Trends in Cognitive Sciences*, 7: 320-24.
- 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.
- Zak, Paul J.** *Moral Markets: The Critical Role of Values in the Economy*. Princeton and Oxford: Princeton University Press, 2008.

## 7. Appendix: Protected-Values survey

*(To conserve Journal space, this appendix can be made available online.)*

Note: This is a translation. The original Protected-Values survey was conducted in German.

Because CEOs' compensation levels depend on the earnings they report to their shareholders, CEOs have an incentive to modify reports to shareholders. What is your opinion on CEOs modifying company information in reports?

Please choose the appropriate category. This is:

|                         |   |   |   |   |   |   |   |                   |
|-------------------------|---|---|---|---|---|---|---|-------------------|
| Very immoral            | 1 | 2 | 3 | 4 | 5 | 6 | 7 | very moral        |
| Not at all praiseworthy | 1 | 2 | 3 | 4 | 5 | 6 | 7 | very praiseworthy |
| Not at all blameworthy  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | very blameworthy  |
| Not at all outrageous   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | very outrageous   |
| Not at all acceptable   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | very acceptable   |

CEOs have an opportunity to modify information in the reports they provide to their shareholders. Some view such modification as a violation of truthfulness, others regard it as acceptable protection of personal interests. What do you think about the value of truthfulness in such a situation?

Truthfulness is about something

|   |   |   |   |   |   |   |   |                |
|---|---|---|---|---|---|---|---|----------------|
| ... that one should not sacrifice, no matter what the (material or other) benefits. |   |   |   |   |   |   |   |                |
| strongly disagree   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | strongly agree |
| ... for which I think it is right to make a cost-benefit analysis.                  |   |   |   |   |   |   |   |                |
| strongly disagree   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | strongly agree |
| ... that cannot be measured in monetary terms.                                      |   |   |   |   |   |   |   |                |
| strongly disagree   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | strongly agree |
| ... about which I can be flexible if the situation demands it.                      |   |   |   |   |   |   |   |                |
| strongly disagree   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | strongly agree |

# 8. Tables and Figures

**Table I: Manipulation checks**

This table presents the results from the manipulation checks in the experiment. 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 appropriate reordering (the direction of the scale varied between questions), a value of +2 indicates that the action was seen as honest, non-manipulative, associated with a personal financial loss, and based on a long-term view. The table indicates that announcing 31 (35) cents was seen as honest (dishonest), non-manipulative (manipulative), implying a personal financial loss (gain), and relating to a long-term (short-term) view.

|  | 31 cents | 35 cents | t-test |
|--|----------|----------|--------|
| <b>Honest vs. Dishonest</b>                                | 1.61     | -1.17    | 26.47  |
| <b>Non-manipulative vs. Manipulative</b>                   | 1.39     | -1.14    | 22.70  |
| <b>Personal financial loss vs. Personal financial gain</b> | 0.93     | -1.25    | 19.50  |
| <b>Long-term view vs. Short-term view</b>                  | 0.99     | -1.1     | 18.53  |

**Table II: Descriptive statistics (next page)**

**Table III: Behavior across phases and costs of truth-telling**

This table presents the percentages of participants announcing 31 cents of earnings per share across the various cost conditions and phases of the experiment.

| Costs                                   |   | Phase 1 | Phase 2 (norms) |          |         | All           |
|---|---|---------|-----------------|----------|---------|---------------|
|   |   |         | Positive        | Negative | Control |               |
| Percent of subjects announcing 31 cents |   |         |                 |          |         |               |
| lowest                                  | 1 | 81.99%  | 63.10%          | 80.68%   | 71.91%  | 77.01%        |
|   | 2 | 52.11%  | 33.33%          | 65.91%   | 49.44%  | 50.96%        |
|   | 3 | 31.42%  | 15.48%          | 54.55%   | 33.71%  | 33.14%        |
|   | 4 | 22.99%  | 15.48%          | 35.23%   | 26.97%  | 24.52%        |
| highest                                 | 5 | 21.07%  | 14.29%          | 31.82%   | 23.60%  | 22.22%        |
| Total                                   |   | 41.92%  | 28.33%          | 53.64%   | 41.12%  | <b>41.57%</b> |

**Table II: Descriptive statistics**

This table presents descriptive statistics for the experiment. *TRUTHFUL CHOICE* is equal to 1 if a participant chose to announce earnings of 31 cents, while it is equal to 0 if the participant chose to announce 35 cents. *EFFORT CHOICE* is equal to 1 if an individual chose to do five calculations, while it is equal to 0 if the participant chose to do one calculation. *COST* is equal to 1, 2, 3, 4, and 5, when the costs of truth-telling are CHF 0, CHF 0.30, CHF 0.60, CHF 0.90, and CHF 1.20, respectively. *PHASE 1* is equal to 1 in the first phase of the experiment, and 0 otherwise. *CONTROL* is equal to 1 for the second phase with no additional information, and 0 otherwise. *POSITIVE NORM* is equal to 1 for the second phase with the positive social norm (high acceptance of earnings management), and 0 otherwise. *NEGATIVE NORM* is equal to 1 for the second phase with the negative social norm (disapproval of earnings management), and 0 otherwise. *SEX* is equal to 1 for female participants, and 0 for male participants. *AGE* is the age in years. *PSYCHOLOGY* is equal to 1 for psychology participants (“psychologists”), and 0 otherwise. *OTHER* is equal to 1 for participants from fields other than psychology and economics, and 0 otherwise. *MARGINAL UTILITY* is equal to 5 (1) for participants who declared that they would benefit “a lot more” (“a lot less”) than their peers from receiving CHF 50. (For details, see the text). *SELFDECEIT* and *EXTDECEIT* are measures of the tendency to give socially acceptable or desirable responses; the participants who exhibited more socially desirable responses scored higher on both scales. Participants were provided with an item and asked, on a 5-point scale ranging from -2 to +2, the extent to which they judged announcing 35 cents as hurting other stakeholders. The corresponding variable is called *35HURTS*. *PROTECTED VALUES (PV)* is the average of all nine items of the survey described in the text. Higher numbers correspond to stronger protected values.

| Variable   | Obs  | Mean  | Std. Dev. | Min  | Max |     |
|--|------|-------|-----------|------|-----|-----|
| <b>Choice variables</b>                            |      |       |           |      |     |     |
| <i>TRUTHFUL CHOICE</i>                             | 2610 | 0.42  | 0.49      | 0    | 1   |     |
| <i>EFFORT CHOICE</i>                               | 2610 | 0.76  | 0.43      | 0    | 1   |     |
| <b>Conditions</b>                                  |      |       |           |      |     |     |
| <i>COST</i> (of no earnings management)            | 2610 | 3.00  | 1.41      | 1    | 5   |     |
| <i>PHASE 1</i>                                     | 2610 | 0.50  | 0.50      | 0    | 1   |     |
| <i>CONTROL GROUP PHASE 2</i>                       | 2610 | 0.17  | 0.38      | 0    | 1   |     |
| <i>POSITIVE NORM PHASE 2</i>                       | 2610 | 0.16  | 0.37      | 0    | 1   |     |
| <i>NEGATIVE NORM PHASE 2</i>                       | 2610 | 0.17  | 0.37      | 0    | 1   |     |
| <b>Individual-level control variables</b>          |      |       |           |      |     |     |
| <i>SEX</i>   | 261  | 0.42  | 0.49      | 0    | 1   |     |
| <i>AGE</i>   | 261  | 24.38 | 4.52      | 18   | 53  |     |
| <i>MARGINAL UTILITY</i>                            | 261  | 3.26  | 0.80      | 1    | 5   |     |
| <i>35HURTS</i> (35 cents seen as hurting others)   | 261  | 0.74  | 1.07      | -2   | 2   |     |
| <i>SELFDECEIT</i>                                  | 261  | 44.73 | 8.29      | 17   | 70  |     |
| <i>EXTDECEIT</i>                                   | 261  | 34.55 | 9.28      | 11   | 57  |     |
| <b>Protected values</b>                            |      |       |           |      |     |     |
| PROTECTED VALUES ( <i>PV</i> ) in experiment       | 261  | 4.82  | 1.03      | 1    | 7   |     |
| PROTECTED VALUES in separate survey                | 123  | 4.83  | 1.03      | 1.67 | 7   |     |
|  | 1%   | 10%   | 25%       | 75%  | 90% | 99% |
| Percentiles of PROTECTED VALUES in experiment      | 2.6  | 3.6   | 4         | 5.6  | 6.1 | 7   |
| Percentiles of PROTECTED VALUES in separate survey | 2.2  | 3.5   | 3.9       | 5.3  | 6.1 | 7   |

**Table IV: Earnings management and protected values**

This table presents coefficients of logit regressions. The dependent variable is *TRUTHFUL CHOICE*, which is equal to 1 when a participant chose to announce 31 cents and equal to 0 otherwise. The explanatory variables are defined in the text and in the notes to Table III. Robust standard errors, obtained by clustering at the individual level, appear in parentheses below coefficient estimates. \*\*\*, \*\*, and \* indicate significance at the 1%, 5%, and 10% levels, respectively.

|                                | (1)                 | (2)                 | (3)                   | (4)                   | (5)                   |
|--------------------------------|---------------------|---------------------|-----------------------|-----------------------|-----------------------|
|                                | Full sample         | Full sample         | Phase 1               | Phase 2               | Full sample           |
| Protected values ( <i>PV</i> ) | 0.708***<br>(0.101) | 0.686***<br>(0.108) | 0.822***<br>(0.140)   | 0.891***<br>(0.147)   | 0.854***<br>(0.136)   |
| Sex                            |                     | 0.260<br>(0.211)    | 0.311<br>(0.274)      | 0.209<br>(0.273)      | 0.259<br>(0.254)      |
| Age                            |                     | -0.0230<br>(0.0237) | -0.0221<br>(0.0312)   | -0.0382<br>(0.0319)   | -0.0301<br>(0.0306)   |
| Psychology                     |                     | 0.0471<br>(0.244)   | 0.0186<br>(0.310)     | 0.166<br>(0.319)      | 0.0926<br>(0.295)     |
| Other studies                  |                     | 0.107<br>(0.298)    | 0.246<br>(0.397)      | 0.0637<br>(0.414)     | 0.155<br>(0.378)      |
| COST of no earnings management |                     |                     | -0.816***<br>(0.0588) | -0.673***<br>(0.0579) | -0.744***<br>(0.0531) |
| Positive norm                  |                     |                     |                       | -0.532*<br>(0.285)    | -0.693***<br>(0.194)  |
| Negative norm                  |                     |                     |                       | 0.753***<br>(0.277)   | 0.622***<br>(0.174)   |
| Control group                  |                     |                     |                       |                       | -0.147<br>(0.179)     |
| Observations                   | 2610                | 2610                | 1305                  | 1305                  | 2610                  |

**Table V: Differential resistance to economic costs and social norms**

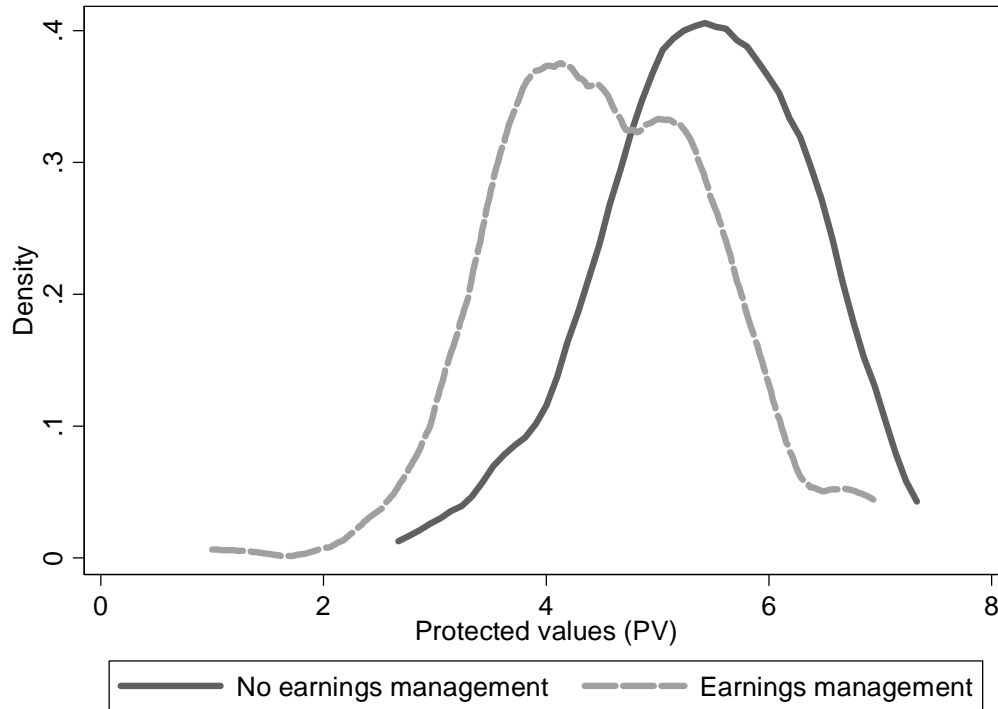
The top part of this table presents coefficients of logit regressions. The dependent variable is *TRUTHFUL CHOICE*, which is equal to 1 when a participant chose to announce 31 cents and equal to 0 otherwise. Column 1 uses the full sample. Regressions in the remaining columns are calculated separately for the participants in the quantile described at the top of the respective column. The explanatory variables are defined in the text and in the notes to Table III. Robust standard errors, obtained by clustering at the individual level, appear in parentheses below coefficient estimates. The bottom part of this table provides differences between coefficient estimates for the top and bottom quantile relevant in the respective column, for the four variables listed in each row. The z-statistics for the significance of these differences are in parentheses. As we have independent samples, these statistics are computed as  $(\beta_i - \beta_j) / (\sqrt{se(\beta_i)^2 + se(\beta_j)^2})$  for two quantiles  $i, j$ . For the quartile and quintile only the coefficients for the top quantiles are shown. In computing the test statistics, the corresponding (untabulated) coefficients and standard errors from the bottom quartile and quintile *PV* groups are also used. \*\*\*, \*\*, and \* indicate significance at the 1%, 5%, and 10% levels, respectively.

|  | (1)<br>Full sample    | (2)<br>Bottom third<br><i>PV</i> | (3)<br>Middle third<br><i>PV</i> | (4)<br>Top third<br><i>PV</i> | (5)<br>Top quartile<br><i>PV</i> | (6)<br>Top quintile<br><i>PV</i> |
|--|-----------------------|----------------------------------|----------------------------------|-------------------------------|----------------------------------|----------------------------------|
| Protected values ( <i>PV</i> )   | 0.869***<br>(0.141)   |                                  |                                  |                               |                                  |                                  |
| Sex  | 0.258<br>(0.258)      | -0.101<br>(0.388)                | 1.187***<br>(0.359)              | 0.245<br>(0.527)              | 0.337<br>(0.739)                 | -0.651<br>(0.959)                |
| Age  | -0.0273<br>(0.0292)   | -0.0178<br>(0.0559)              | 0.0331<br>(0.0434)               | -0.0498<br>(0.0397)           | -0.0229<br>(0.0566)              | -0.0431<br>(0.0598)              |
| Psychology   | 0.119<br>(0.299)      | 0.460<br>(0.478)                 | -1.337***<br>(0.487)             | 0.626<br>(0.555)              | 0.595<br>(0.747)                 | 1.767*<br>(0.981)                |
| Other studies  | 0.206<br>(0.378)      | 0.369<br>(0.327)                 | -0.895<br>(0.650)                | 0.747<br>(0.643)              | 0.645<br>(0.737)                 | 0.934<br>(0.796)                 |
| <i>COST</i> of no earnings management  | -0.802***<br>(0.0618) | -1.065***<br>(0.151)             | -0.776***<br>(0.0963)            | -0.527***<br>(0.0634)         | -0.437***<br>(0.0669)            | -0.433***<br>(0.0802)            |
| Positive norm  | -0.766***<br>(0.194)  | -1.178***<br>(0.305)             | -0.546*<br>(0.308)               | -0.362<br>(0.382)             | -0.122<br>(0.408)                | 0.263<br>(0.550)                 |
| Negative norm  | 0.658***<br>(0.178)   | 0.991***<br>(0.338)              | 0.360<br>(0.280)                 | 0.532*<br>(0.290)             | 0.0235<br>(0.334)                | -0.222<br>(0.364)                |
| Control group  | -0.188<br>(0.181)     | -0.431<br>(0.390)                | -0.256<br>(0.298)                | 0.0545<br>(0.283)             | 0.199<br>(0.339)                 | 0.136<br>(0.360)                 |
| <i>PV</i> * <i>COST</i>  | 0.245***<br>(0.0539)  |                                  |                                  |                               |                                  |                                  |
| <i>PV</i> * Positive norm  | 0.399**<br>(0.200)    |                                  |                                  |                               |                                  |                                  |
| <i>PV</i> * Negative norm  | -0.169<br>(0.210)     |                                  |                                  |                               |                                  |                                  |
| <i>PV</i> * Control group  | 0.139<br>(0.190)      |                                  |                                  |                               |                                  |                                  |
| Observations   | 2610                  | 920                              | 870                              | 820                           | 600                              | 490                              |
| Number of participants   | 261                   | 92                               | 87                               | 82                            | 60                               | 49                               |
| <b>Differences of coefficients between the top and bottom quantiles and z-statistics for significance of differences</b> |                       |                                  |                                  |                               |                                  |                                  |
| <i>COST</i> of no earnings management  |                       |                                  |                                  | 0.538***<br>(3.29)            | 0.91***<br>(5.01)                | 1.157***<br>(5.54)               |
| Positive norm  |                       |                                  |                                  | 0.816*<br>(1.67)              | 1.002*<br>(1.84)                 | 1.403**<br>(2.24)                |
| Negative norm  |                       |                                  |                                  | -0.459<br>(1.03)              | -0.746<br>(1.39)                 | -1.575**<br>(2.42)               |
| Control group  |                       |                                  |                                  | 0.486<br>(1.01)               | 0.962<br>(1.50)                  | 0.335<br>(0.51)                  |

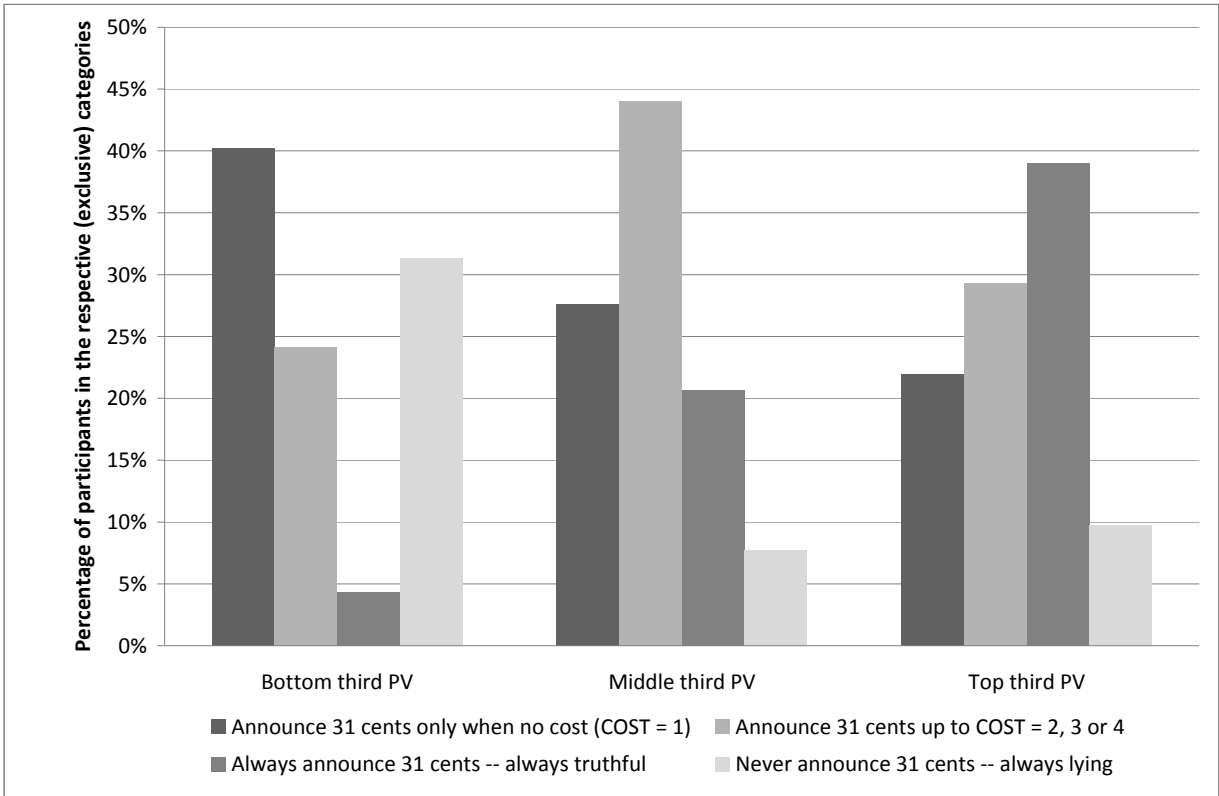
**Table VI: Earnings management and protected values, further results**

Columns 1 – 4 of this table present coefficients of logit regressions. The dependent variable is *TRUTHFUL CHOICE*, which is equal to 1 when a participant chose to announce 31 cents and equal to 0 otherwise. The explanatory variables are defined in the text and in the notes to Tables III and IV. To conserve space, the table only presents the coefficients of the key variables of interest, but all regressions include the usual controls. Robust standard errors, obtained by clustering at the individual level, appear in parentheses below coefficient estimates. Column 5 presents the differences between coefficient estimates for the top quartile *PV* and the bottom quartile *PV* groups, for the variables listed in each row. z-statistics for the significance of the differences appear below the differences. The test statistics are calculated as described in Table V. \*\*\*, \*\*, and \* indicate significance at the 1%, 5%, and 10% levels, respectively.

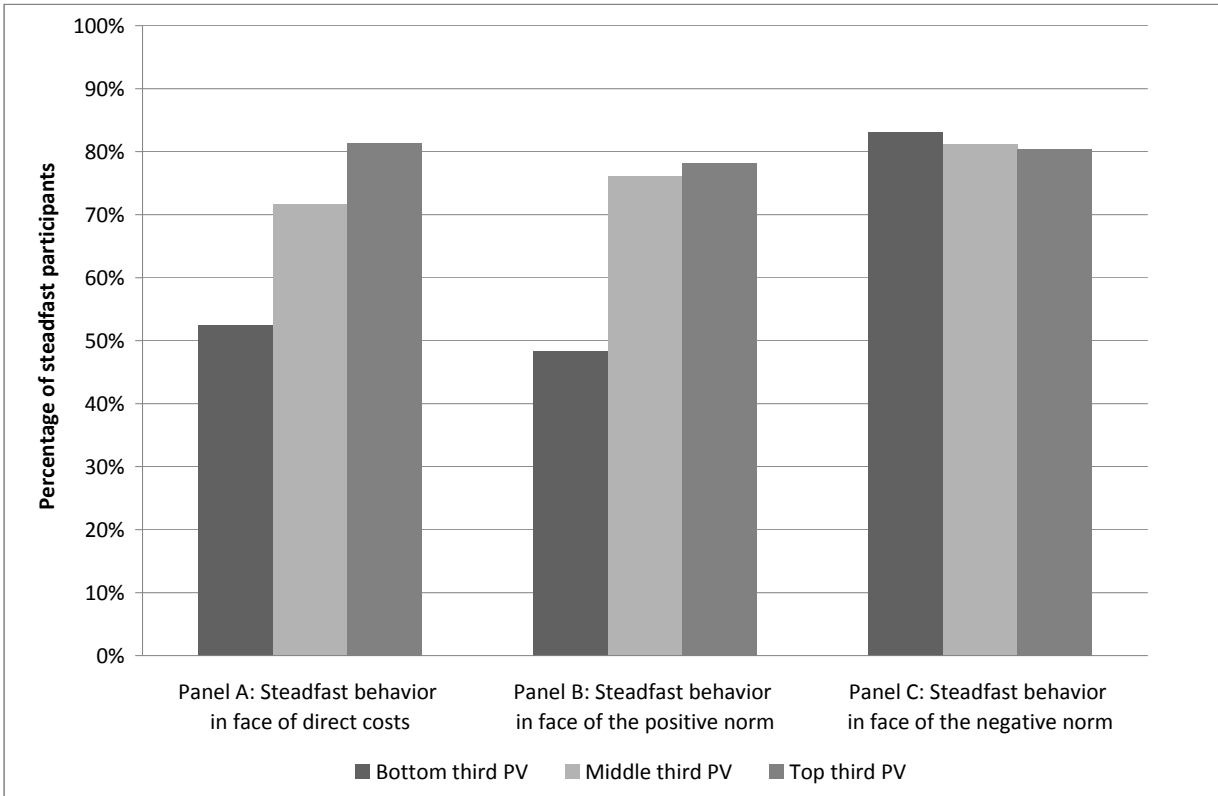
|   | (1)<br>Bottom<br>quarter <i>PV</i> | (2)<br>Second<br>quarter <i>PV</i> | (3)<br>Third<br>quarter <i>PV</i> | (4)<br>Top quarter<br><i>PV</i> | (5)<br>Difference<br>Top - Bottom |
|---|------------------------------------|------------------------------------|-----------------------------------|---------------------------------|-----------------------------------|
| <b>Panel A</b>  |                                    |                                    |                                   |                                 |                                   |
| <i>COST</i> of no earnings management                               | -1.375***<br>(0.171)               | -0.863***<br>(0.133)               | -0.781***<br>(0.0977)             | -0.443***<br>(0.0689)           | 0.932***<br>(5.05)                |
| Positive norm   | -1.159***<br>(0.360)               | -0.948**<br>(0.385)                | -0.831**<br>(0.374)               | -0.153<br>(0.410)               | 1.006*<br>(1.84)                  |
| Negative norm   | 0.777*<br>(0.429)                  | 0.847**<br>(0.360)                 | 0.938***<br>(0.328)               | 0.0677<br>(0.340)               | -0.709<br>(1.30)                  |
| Self-reported marginal utility                                      | -0.415**<br>(0.201)                | 0.238<br>(0.386)                   | 0.313<br>(0.365)                  | -0.319<br>(0.278)               | 0.096<br>(0.28)                   |
| <b>Panel B</b>  |                                    |                                    |                                   |                                 |                                   |
| <i>COST</i> of no earnings management                               | -1.347***<br>(0.169)               | -0.871***<br>(0.135)               | -0.847***<br>(0.101)              | -0.442***<br>(0.0680)           | 0.905***<br>(4.97)                |
| Positive norm   | -1.127***<br>(0.359)               | -0.883**<br>(0.380)                | -0.880**<br>(0.392)               | -0.0806<br>(0.401)              | 1.046**<br>(1.94)                 |
| Negative norm   | 0.762*<br>(0.422)                  | 0.767**<br>(0.389)                 | 1.040***<br>(0.332)               | -0.0290<br>(0.342)              | -0.791<br>(1.46)                  |
| <i>35HURTS</i>  | -0.0284<br>(0.225)                 | -0.287<br>(0.217)                  | -0.665***<br>(0.177)              | 0.254<br>(0.266)                | 0.282<br>(0.81)                   |
| <b>Panel C</b>  |                                    |                                    |                                   |                                 |                                   |
| <i>COST</i> of no earnings management                               | -1.366***<br>(0.173)               | -0.875***<br>(0.132)               | -0.776***<br>(0.0993)             | -0.438***<br>(0.0683)           | 0.898***<br>(4.83)                |
| Positive norm   | -1.085***<br>(0.374)               | -0.869**<br>(0.388)                | -0.893**<br>(0.373)               | -0.151<br>(0.413)               | 0.934*<br>(1.68)                  |
| Negative norm   | 0.793*<br>(0.420)                  | 0.820**<br>(0.358)                 | 0.980***<br>(0.326)               | 0.0426<br>(0.338)               | -0.750<br>(1.39)                  |
| Self deceit   | 0.0201<br>(0.0181)                 | 0.0398<br>(0.0261)                 | 0.00885<br>(0.0251)               | 0.00479<br>(0.0256)             | -0.0153<br>(0.49)                 |
| External deceit   | 0.0317<br>(0.0234)                 | -0.000690<br>(0.0308)              | -0.00572<br>(0.0339)              | -0.0138<br>(0.0262)             | -0.046<br>(1.31)                  |
| All regressions: Control group, Sex, Age, Psychology, Other studies | Yes                                | Yes                                | Yes                               | Yes                             |                                   |
| Observations  | 690                                | 640                                | 680                               | 600                             |                                   |



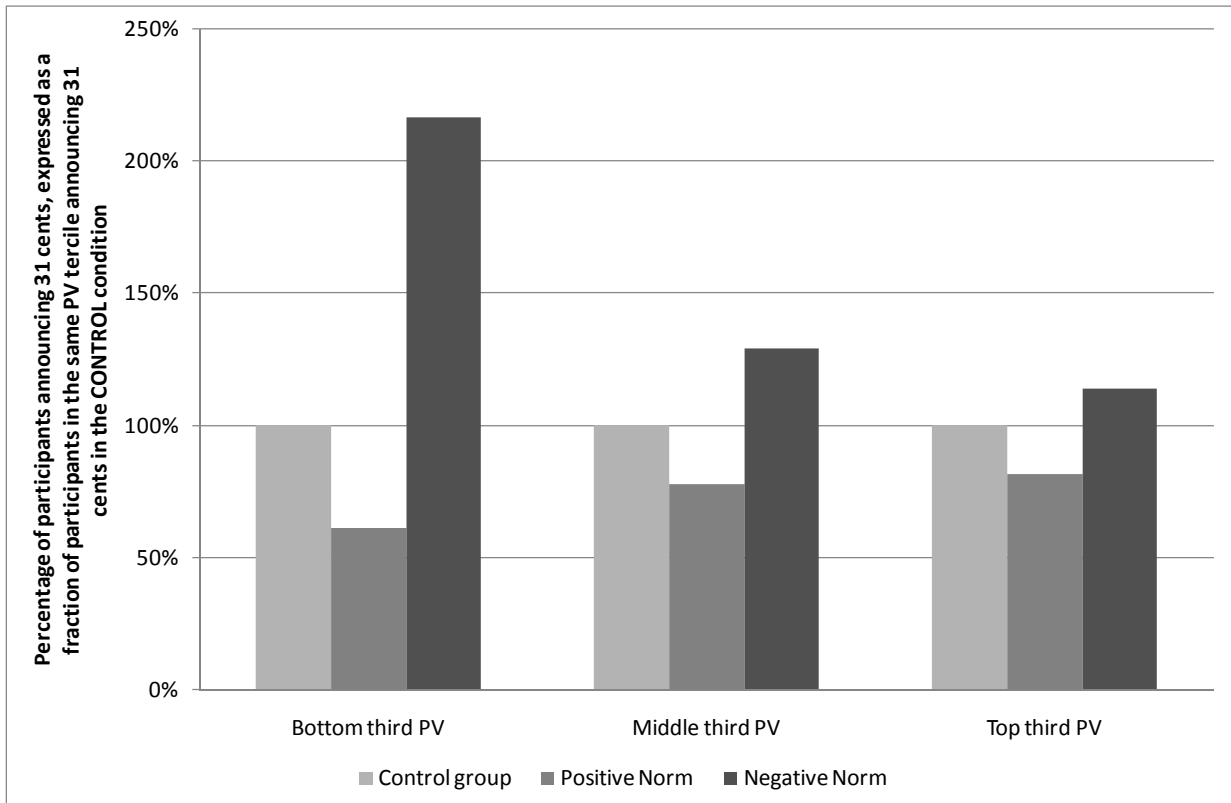
**Figure 1:** This figure plots the distribution of the strength of protected values of the participants who chose to manage earnings (broken line) and for those participants who willingly bore economic costs for not managing earnings (solid line). The figure is plotted for the median cost level in Phase 1.



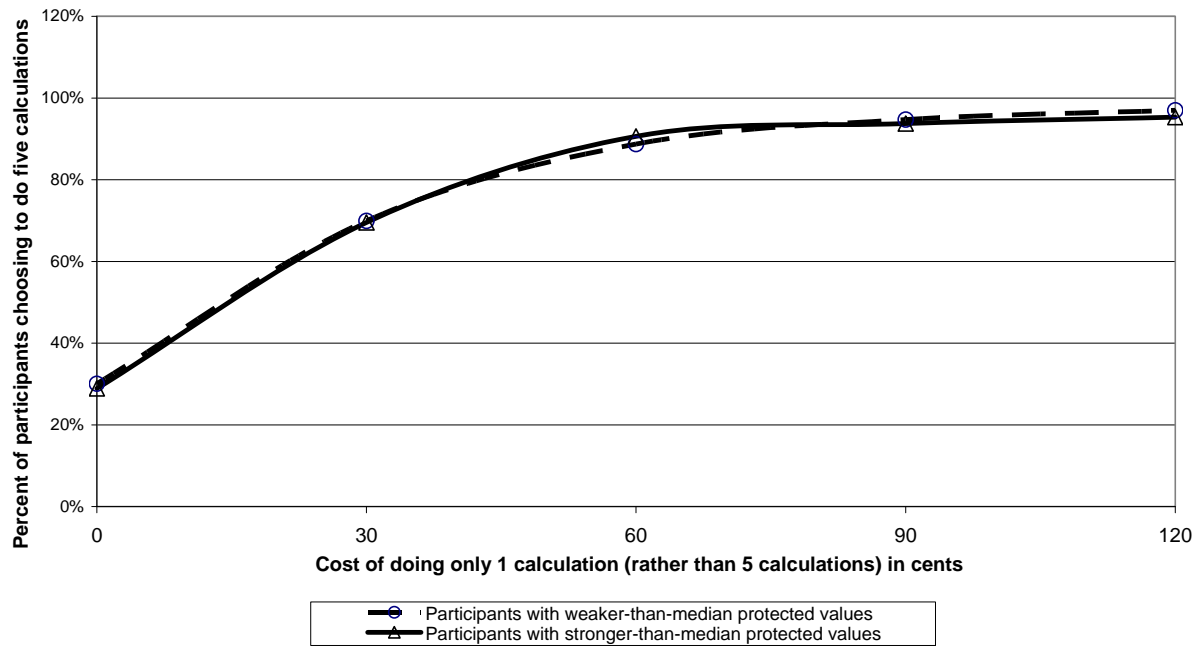
**Figure 2:** This figure plots the percentages of participants whose Phase 1 behavior can be described by the four categories of choices shown, respectively, split up according to their *PV* levels. The second column from the left in each panel shows the sum of the percentage of participants announcing 31 cents at both and only both *COST* levels 1 and 2, plus the percentage of participants announcing 31 cents at all three and only all three *COST* levels 1, 2, and 3, plus the percentage of participants announcing 31 cents at all four and only all four *COST* levels 1, 2, 3, and 4. To simplify exposition, the figure omits the roughly 1% of cases where participants were not consistent in their choices, that is, they announced 35 cents at a low cost level, but 31 cents at a high cost level.



**Figure 3:** This figure plots the percentage of steadfast participants, split up into three terciles of *PV*. Panel A shows the percentage of those participants who announced 31 cents at the previous cost level who also announced 31 cents at the next cost level, averaged over all cost levels (using only data from Phase 1). For example, among those in the highest tercile of *PV*, 90.2% reported the truth when it was free, 68.3% (=75.7% of 90.2%) reported the truth at the next cost level, 52.4% (=76.8% of 68.3%) at the next cost level, 42.7% (=81.4% of 52.4%) at the next cost level, and 39% (=91.4% of 52.4%) at the highest cost level. On average, therefore, 81.3%  $(=(1/4)*(75.7%+76.8%+81.4%+91.4%))$  exhibited steadfast behavior in face of direct economic costs of truthfulness. Panel B plots the percentage of those participants who announced 31 cents in Phase 1 who announced 31 cents also under the positive norm in Phase 2 at the same cost level, jointly considering all cost levels. (That is, we added, over all *COST* levels, the number of participants who remained truthful and divided the resulting sum by the sum of the number of participants who had reported the truth in Phase 1 and went through the positive norm manipulation. The picture looks similar if one calculates the respective percentages at each *COST* level and then averages them.) Panel C shows the percentage of those participants who announced 35 cents in Phase 1 who announced 35 cents also under the negative norm in Phase 2 at the same cost level, jointly considering all cost levels. (We are interested in the imperviousness of participants. Alternatively, we could plot the percentage of those who switch, i.e., who are influenced by the negative norm. This is just 1 minus the number shown.)



**Figure 4:** This figure shows, for each of the three PV terciles, the percentage of participants who announced 31 cents as a fraction of those who announced 31 cents in the *CONTROL* condition. The figure uses all *COST* situations. Averaging over all *COST* situations, in the lowest, middle, and top terciles, 17.9%, 42.1%, and 60.6%, respectively, announced 31 cents in the *CONTROL* condition.



**Figure 5:** This figure plots the percentage of participants choosing to do five calculations distributed over the cost of doing only one calculation. The solid line corresponds to participants with stronger-than-median protected values, while the dashed line corresponds to participants with weaker-than-median protected values.