Dhami, Sanjit

Working Paper

Human Ethics and Virtues: Rethinking the Homo-Economicus Model

CESifo Working Paper, No. 6836

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Dhami, Sanjit (2017) : Human Ethics and Virtues: Rethinking the Homo-Economicus Model, CESifo Working Paper, No. 6836, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:
https://hdl.handle.net/10419/174959

CESifo WORKING PAPERS

6836
2017

December 2017

Human Ethics and Virtues: Rethinking the Homo-Economicus Model

Sanjit Dhami

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editors: Clemens Fuest, Oliver Falck, Jasmin Gröschl
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 6836*
Category 13: Behavioural Economics

Human Ethics and Virtues: Rethinking the Homo-Economicus Model

Abstract

The neoclassical model in economics envisages humans as amoral and self-regarding (Econs). This model, also known as the homo-economicus model, is not consistent with the empirical evidence. In light of the evidence, the continued use of the homo-economicus model is baffling. It also stymies progress in the field by putting the burden of adjustment on auxiliary assumptions that need to compensate for an unrealistic picture of human motivation and behavior. This essay briefly outlines the evidence for a more inclusive picture of humans in which ethics and morality play a central role. It argues for replacing the homo-economicus model with a homo-behavioralis model that has already enabled great progress to be made in the field of behavioral economics.

JEL-Codes: D900, D640.

Keywords: ethics, morality, intrinsic motivation, consequentialistic choices, lying-aversion, guilt-aversion, markets and morality, moral balancing, self-image, self-serving justifications, partial lying, third party punishment, delegation, social identity, moral suasion.

Sanjit Dhami
Division of Economics
University of Leicester
University Road
United Kingdom – LeicesterLE1 7RH
sd106@le.ac.uk

21 December 2017
This is a longer version of an article to be published in an abridged form in the Handbook of Ethics and Economics. Oxford University Press: Oxford. I am grateful for comments to Herbert Gintis, Björn Bartling, Gary Charness, Marie Clarie Villeval, and Smruti Bulsari.

1. Introduction

The dominant paradigm in economics, *neoclassical economics*, is based on the *homo-economicus* model. Fictional analogues of humans in this model, *Econs*, are assumed to be *amoral* and entirely *self-regarding*, devoid of any intrinsic morality, or any desire to behave in an ethical manner. Econs do not exhibit an *intrinsic preference* for honesty; truth-telling; keeping promises; trusting others and being trustworthy; reciprocating kind and unkind behavior of others; and caring about the fairness of procedures. Econs also have no feelings of remorse or guilt from letting down the expectations of others. Econs strive solely to maximize their own material well-being (*self-regarding* preferences) without regard to the well-being of others, i.e., they lack *other-regarding preferences*.[1]

Contemporary research and teaching in economics continues to be based, almost exclusively, on analyzing the behavior of Econs. The validity of this model is often taken as an article of faith among economists. Social scientists working in other disciplines are likely to be staggered by the widespread acceptance of such a worldview among economists. However, when pressed, many economists might express the view that they do not believe in the 'literal truth' of such a model, but that such a model provides a 'good approximation' to the real world. However, the 'good approximation' part of the argument is never formally demonstrated, merely asserted.

I suspect that a sizeable number of economists are likely to argue that giving up a worldview based entirely on Econs will squander the hard-earned discipline of economic models and will open the way for a proliferation of ad-hoc models. This is a common, but deeply flawed argument. Discipline, and progress, in science arises from building models that are in conformity with the empirical evidence and being prepared to amend models as new evidence emerges. As Richard Feynman, the Nobel Prize winner in physics is once reported to have said: "We are trying to prove ourselves wrong as quickly as possible, because only in that way can we find progress." If a model based on Econs is rejected, then alternative models that might be in better conformity with the evidence need not be ad-hoc.[2]

This essay will argue that the homo-economicus model is not consistent with the empirical evidence. I will restrict myself to discussing issues of *morality* and *ethics* in this paper, with particular emphasis on the incentive to lie. Readers interested in the empirical validity of the self-regarding assumption behind the homo-economicus model and the evidence for other-regarding preferences can consult several good sources. In this survey, I omit this discussion; see, for instance, Camerer (2003), Fehr and Schmidt (2006), Dhami (2016, Part 2), and Gintis (2017).[3]

[1] To be sure, neoclassical economics can be amended to include some forms of other-regarding preferences, such as "keeping up with the Joneses" or engaging in "snob or conspicuous consumption". However, these features play, at best, a peripheral role in the *actual practice* of neoclassical economics, which is the benchmark that we are interested in.

[2] For a more detailed discussion of these issues, see the introductory chapter in Dhami (2016). This is likely to be particularly valuable for those economists who believe in various homegrown scientific methods that have no analogues in any of the other successful sciences or in the philosophy of science.

[3] Gintis (2017, p. 50) suggests useful terminology on the two-way classification between preferences and ethicality. In my survey, I shall abstract from the middle column in his Figure 3.1 corresponding

There should be no presumption that intrinsic human morality and human virtues dilute the rationality assumption in economics. Rationality simply requires that people should have consistent preferences. The presence of other-regarding preferences still leads to rational choices in this sense (Andreoni and Miller, 2002). This essay outlines the emerging evidence on ethics and morality, as well as some of the emerging theoretical insights, in behavioral economics.

Neoclassical economics does not deny the existence of human behavior such as reciprocity, truth-telling, and keeping promises. However, it ascribes the cause of this behavior to *extrinsic preferences* for maximizing one's own material well-being. Econs will choose to tell the truth when the current benefits from lying are lower than the discounted future costs of such lies. In strategic situations, Econs might be induced to reveal the truth because extremely clever contracts guarantee lower payoffs from lying, relative to truth-telling; this approach lies at the heart of *contract theory* and *mechanism design*. The motivation to be intrinsically honest is not taken into account in these contracts because Econs are amoral. Rather, virtuous behavior is assumed to be merely instrumental in increasing material well-being. Fortunately, a strength of the homo-economicus model is that it typically makes precise, testable, predictions, so we can test whether morality is intrinsic or extrinsic and whether morality is influenced by factors such as the context, frame, the size of incentives, and competition. I shall argue below that the predictions arising from this framework are not well supported by the evidence.

Several decades of research in behavioral economics has shown that if the aim is to better explain and understand human behavior, then *homo-behavioralis* is a superior candidate to replace homo-economicus. Homo-behavioralis cares for material interests and for extrinsic incentives (just like homo-economicus), but also exhibits a strong sense of morality, considers the ethicality of alternative options, responds to intrinsic incentives, and is conditionally reciprocal. Homo-behavioralis has been central to the development of behavioral and experimental economics–the fastest growing and most exciting development in economics in recent decades (Camerer, 2003; Dhami, 2016, Gintis, 2017).

The empirical evidence shows that exclusive reliance on the homo-economicus model has arguably reduced the ability of economic theory to make more realistic predictions. It also appears to have had a determinantal effect on other disciplines, such as management, that have borrowed the basic neoclassical model. Gintis and Khurana (2016) express well the frustration with this model when they write: "*Business schools have widely responded to criticism by adding a course on "business ethics" to the MBA curriculum. While welcome, such a move cannot compensate for the generally incorrect and misleading characterization of human motivation, based on the neoclassical Homo economicus perspective, promulgated in courses on managerial behavior and corporate culture. This model must be directly attacked and replaced by a more accurate model of human motivation.*"

This essay is divided into several sections; yet limitations of space dictate that the treatment of each topic is relatively brief. The sections often overlap. For instance, there

to other-regarding preferences and deal with (1) amoral self-regarding preferences, and (2) preferences that demonstrate intrinsic morality (or *universalist preferences* in Gintis's terminology). On the ethicality dimension, Gintis makes a distinction between differences in two types of human persona, private and public. I consider these issues below in Section 7.5.

are separate sections on gender effects and on the field evidence, yet, unavoidably, gender data and field data are often considered in the other sections too. Section 2 explains three different canonical experiments that have been used in the lab to study lying behavior. Section 3 considers the field evidence, while Section 4 considers the external validity of lab evidence. Section 5 considers gender effects on lying. Section 6 examines the effects of incentives on lying. Section 7 explores some of the microfoundations for moral behavior. Section 8 considers many different aspects of moral behavior, such as those arising from gain and loss frames, delegation, third party punishment, moral suasion, and social identity. Section 9 introduces how one might make use of psychological game theory to consider emotions such as guilt, shame, and intentions, that may underpin moral behavior. Section 10 examines the relation between markets and morality. Section 11 considers the cross country evidence on honesty. Section 12 touches upon the some of the neuroeconomic evidence. Finally, 13 concludes.

2. An introduction to experimental methods on lying behavior

An advantage of experiments over field data is that lying behavior can be examined under relatively more controlled conditions that potentially unravels lying behavior either at the individual or aggregate level. Second, subjects in lab experiments can be assured, as far as possible, that their lying will not be observable to a third party, which enables more accurate measurements. For instance, individuals may be asked to roll a die, or toss a coin, in private, and, self-report the outcome, which translates into a pre-determined reward. If despite costless lying, many people choose to remain honest, then the incidence of lying is likely to be even lower when lying is costly. Thus, experiments reveal valuable information that helps us to build a better picture of human ethicality and place bounds on lying behavior.

We begin with some useful terminology (Gneezy, 2005; Erat and Gneezy, 2012). In neoclassical economics, economic agents, or Econs, feel no remorse, disutility, or guilt, in telling *selfish black lies*–these are lies that benefit the liar but potentially harm others. Since neoclassical economics is based on *extrinsic* human morality only, it then proceeds to determine incentive compatible mechanisms for Econs that induce truth-telling, purely in response to extrinsic economic incentives. In contrast, *white lies* benefit others, but they must not decrease the liar's own utility. Examples include a supervisor who writes his report diplomatically to make it more palatable to a failing employee, so as not to undermine the latter's confidence; or a doctor who, unknown to a suffering patient, gives a placebo to ease his suffering. *Altruistic white lies* may harm the liar but benefit others. Finally, *Pareto white lies* improve the utilities of the liar and of others.

The following terminology may also be useful in some cases. Suppose that action A is honest, but action B is a lie, yet both actions lead to the same monetary payoff for an individual, say, individual 1. However, these actions may harm/benefit individual 2. If individual 1 is indifferent between the two actions, then he/she is said to take a *consequentialistic approach*. Econs are consequentialists. However, if individual 1 has a preference between the actions that is influenced by, say, the effect on the utility of individual 2, then he/she takes a *non-consequentialistic approach*. The bulk of the empirical evidence

Treatment	Option	Payoff of	
		Player 1	Player 2
1	A (64%)	5	6
	B (36%)	6	5
2	A (83%)	5	15
	B (17%)	6	5
3	A (48%)	5	15
	B (52%)	15	5

Table 2.1: Payoffs of the players in each treatment in Gneezy (2005).

indicates that the actions of players are non-consequentialistic.

Consider the following example. Gneezy (2005) asked University of Chicago students: Which of the following two lies is more unfair? In the first case, the seller sells a car to a buyer without revealing that it will cost $250 to fix a pump in the car. In the second case, the seller does not reveal that the brakes are faulty, which will also cost $250 to fix. Despite identical monetary costs, a significantly higher percentage of the respondents judged the second scenario (faulty brakes) to be the more unfair of the two lies. Subjects are not merely consequentialists, but they care about the higher potential risk to the buyer in the second case; indeed, their *intrinsic* human morality plays a critical role in judging actions.

We now consider three different canonical games that have been used to uncover the nature of lying behavior in lab experiments. In the subsequent sections, we shall use these games to build a more complete picture of lying behavior.

2.1. Lying behavior in sender-receiver games

In *sender-receiver games*, player 1 sends a message to player 2, who observes the message, and then takes an action that yields payoffs for both players. Here, we are interested in cheap talk games, in which costless messages do not directly influence the payoffs of players, but do influence the beliefs of other players about the states of the world. Thus, messages indirectly influence the actions of the other players, and the payoffs. Consider the following three treatments in Gneezy (2005) in Table 2.1 in a sender-receiver game.

In each treatment, there are two players. Player 1 (the sender) can observe all the payoffs in Table 2.1, and can send two possible mutually inconsistent messages, A and B, to player 2. Message A says: "Option A will earn you more money than option B." Message B says: "Option B will earn you more money than option A." Player 2 always earns more from option A, so message B is a *lie*. In contrast, the sender always receives a higher payoff from option B.

Unlike player 1, player 2 (the receiver) does not observe the payoffs corresponding to each option in Table 2.1. However, player 2 observes the message sent by player 1, and

then chooses either option A or B. Finally, player 2's chosen option is implemented, and the payoffs corresponding to the chosen option in Table 2.1 accrue to both players.[4]

In a world where senders and receivers are Econs (purely self-regarding and amoral), suppose that senders believe that their messages will be trusted by the receivers. Then, the prediction of the model is that the senders should always choose the payoff maximizing message B. Thus, the message should not be believed by the receivers. In contrast, when the beliefs of the senders are elicited in an *incentive compatible elicitation*, 82% of the senders expected the receiver to trust their message.[5] In actual practice, 78% of the receivers trusted the messages that they received, which is quite close to the expectations of the senders.

The incentive to lie for the sender differs in the three treatments. First, the increase in the sender's payoff from telling a lie is 1 for treatments 1 and 2 but it is 10 for treatment 3. Second, the receiver's loss from being lied to is 1 in treatment 1 but 10 in treatments 2 and 3. Both these incentives play an important role in the sender's decision. The percentage of senders who lied in treatments 1, 2 and 3, respectively, is 36, 17 and 52, and these percentages are statistically different. Thus, the extent of lying responds to the private gain from lying and the cost to the other player. However, a potential confound in the experiment is that it cannot cleanly separately the *role of other-regarding preferences* from *aversion to lying.*

2.2. Lying behavior in the *die-in-the-cup* paradigm

Fischbacher and Föllmi-Heusi (2013) introduced a novel method of measuring aggregate lying, the *die-in-the-cup* paradigm, that is able to rule out potentially confounding factors such as altruism and reputational concerns. In this method, individual reports, which can be lies or be truthful, cannot be checked for lying (but see extensions below). However, the distribution of reports across subjects can be compared to the predicted statistical distribution under truth-telling to determine the extent of aggregate lying.

Subjects roll a six-sided die in private, observe the outcome, and make a report of the observation, which can be any number between 1 to 6. For the numbers 1, 2, 3, 4, and 5, the payoff equals the self-reported number (e.g., \$4 when the report gives the number 4). If the claimed number is 6, then the payoff is zero. Since the experimental design ensures complete anonymity, and outcomes are only privately observed, self-regarding and amoral Econs are predicted to report the number that achieves the highest payoff, i.e., 5.

In the predicted statistical distribution of outcomes under truth-telling, each of the six numbers comes up with a probability 1/6 (or 16.7%). Statistical differences between the actual and the predicted distributions indicate the extent of aggregate departures from truthtelling. The authors also used several control treatments with varying stakes; treatments that imposed a negative externality on another player from lying; and varied

[4]Since the messages are costless, the game does not permit a signaling equilibrium to be sustained.

[5]By *incentive compatible* is meant that there is a strictly positive probability that each of the decisions of the players in the game has a real monetary consequence. In this case, if the beliefs of the senders are accurate, then they receive a monetary prize. Incentive compatibility is required of all experiments in economics; this is true of the results reported in the rest of the paper, unless otherwise stated.

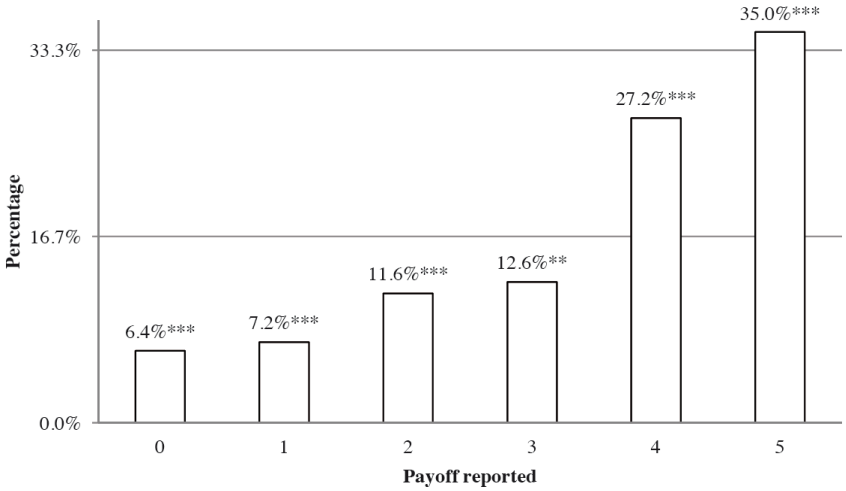


Figure 2.1: Results of the baseline treatment. Source: Fischbacher and Föllmi-Heusi (2013).

the level of anonymity.[6]

In Figure 2.1, based on 389 participants, the horizontal axis shows the reports of the subjects in terms of payoffs (recall, payoffs equal the report for the numbers 1–5 and equal 0 for the number 6). The vertical axis shows the percentage of subjects corresponding to each report; the statistical prediction of 16.7%, if all report truthfully, is also shown. The stars on the top of each percentage sign on a histogram bar denote the significance levels for a two-sided binomial test of the differences between the private reports and the statistical prediction under truthtelling.

The results are as follows. A Kolgomorov-Smirnov test rejects the hypothesis that the distribution of private reports and the predicted distribution under truth-telling are identical. Thus, there is lying at the aggregate level. Numbers 1, 2, 3 and 6 that receive relatively low payoffs, are underreported, while numbers 4, 5 are overreported. Statistically, only 16.7% of the subjects are predicted to get a 6 under truth-telling. Econs should never report a 6, yet 6.4% of the subjects report 6; since $\frac{6.4}{16.7} = 0.38$, the proportion of intrinsically honest subjects is 38.3%. Econs are predicted to only report the outcome 5. Statistically, 16.7% of the subjects would have got a 5 anyway, and the percentage reporting 5 is 35%, hence, the percentage of people who are unethical income maximizers is $\frac{6}{5}(35 - 16.7) = 21.96\%$. Further, 27.2% of the subjects report the number 4, yet only 16.7% are predicted to get a 4 under truth-telling. Hence, some subjects choose to lie but not maximally; these subjects may be termed as *partial liars*, who could be motivated by a desire to maintain a positive self-image, or they suffer costs of lying.

Abeler et al. (2016) conduct a meta analysis of 90 studies that use the Fischbacher-Föllmi-Heusi framework. They found that, on average, subjects forgo three quarters of the potential gain from lying; this could either be due to honest reporting, or due to partial lying. Significantly, they also report that this result is robust to increasing the payoff

[6]In the high anonymity treatment, participants shredded any pieces of paper they had written on and directly took prize money from an envelope without the intervention of the experimenter.

levels, 500 fold relative to a baseline level, to take account of higher stakes.

2.3. Lying in matrix tasks

Another method for detecting lying behavior was introduced by Mazar et al. (2008) in the matrix task in which subjects are given a timed mathematical problem. For instance, detecting the number of pairs of numbers from a given set of numbers in a matrix that add up to 10. Subjects are then given the correct answers and have to self-report the number of their correct answers. Subjects are asked to destroy their answer sheets, and unlike the Fischbacher–Föllmi-Heusi task, there is no objective distribution under truthtelling with which the reported distribution may be compared with. Lying may be discovered in this case if the torn sheets of paper are reconstructed by the experimenter and compared against the claims of the number of self-reported correct answers. In another variant of this method, subjects observe numbers on a computer screen and self-report them. However, unknown to the subjects, the experimenter knows which numbers come up on the screen, which allows lying behavior to be detected. This may be construed by some as, at least, borderline subject deception, a practice that experimental economists are averse to.

3. Evidence from the field

Each form of evidence, lab, field, and survey, has its strengths and weaknesses. Together they build a more complete picture of human ethicality.

Prucker and Sausgruber (2013) conducted field experiments in two towns in Austria, in which a newspaper, costing 0.60 euros, is sold on the street using a booth filled with newspapers. Customers pay into an unmonitored padlocked cashbox, so they can pay a partial amount, the full amount, or not pay at all. There is very little material gain from not paying, and a very small probability of non-payments being discovered.

Two different treatments are run by posting different message in the form of a note on the booth. In the treatment LEGAL, the note reminds customers of the legal norm of paying for the newspaper. In the treatment MORAL, the note gives salience to honest behavior, so it primes customers to follow a social norm. Only a third of the customers pay a strictly positive amount. Average payment in the LEGAL treatment is 0.061 euros and average payment in the MORAL treatment is 0.14 euros. Thus, when subjects are primed to consider a moral norm, self-interest is reduced somewhat in favour of a social norm of honesty. Further, when the moral reminder is removed, it continues to have a positive effect.

In a field study in Germany, conducted by Abeler et al. (2014), participants anonymously tossed a coin in the privacy of their homes, and report the result on the phone. Subjects received a payoff of 15 Euros if they report tails. Econs should only report tails. Although the experiment cannot uncover individual lying, we can compare the reported distribution of tails to the statistical distribution under truth-telling, which predicts 50% tails. Almost all subjects in the field experiment report the truth; indeed the percentage reporting tails is lower than 50%, despite the substantial inducement to lie. These results, like those of Fischbacher–Föllmi-Heusi, are able to rule out potentially confounding

factors such as altruism and reputational concerns. Thus, the results are consistent with a moral cost of lying. The relatively greater dishonesty in comparable lab experiments suggests that either subjects perceive a lower moral cost of lying in the lab, or perhaps that student subjects are predisposed to lie more in such tasks relative to the non-student population. The authors conjecture that lying in a home environment, where one lives with one's family, may be considered more unethical and may violate norms of honest behavior at home.

Utikal and Fischbacher (2013) apply the Fischbacher–Föllmi-Heusi experiments to a sample of female Franciscan nuns in Germany and compare the results with a student subject pool. Students overreport in a manner consistent with the results of Fischbacher–Föllmi-Heusi. However, nuns tell *disadvantageous lies* that harm them but benefit others (altruistic white lies). In terms of Figure 2.1, the percentage of nuns who report the numbers 6, 1, 2, 3 is, respectively, 17, 33, 17, 33. No nuns report the numbers 4 and 5. It could be that nuns who observed the numbers 4 and 5 might have believed that (honestly) reporting these numbers might invite a suspicion of dishonesty. Admittingly, nuns are a very special subject pool, so one cannot generalize these interesting results too much.

Azar et al. (2013) give extra change to clients at an Israeli restaurant and observe if the clients would return the change. They find that long-term clients at the restaurant are more likely to return the extra change, although a majority of the clients (128 out of 192) do not return the change. The authors conjecture that long-time clients may feel guilty cheating the waiter, who they might know, of extra change. Since they may anticipate coming back to the restaurant in the future, they might also worry about their potential reputation with the restaurant staff. One-time customers almost never return the extra change. When the amount of extra change is increased, surprisingly, the authors find that it is more likely to be returned, although the marginal gain, for an Econ, from not returning it, is higher. The authors conjecture that a greater psychological cost must be paid to keep a larger amount of extra change. Females, particularly those who are repeat customers, are more likely to return the extra change. However, an increase in the number of people at same table, has no effect on the extra change returned. A potential confounding factor is that subjects might not notice small amounts of excess change, preventing the option of returning this change.

In Hanna and Wang (2017), 699 students across 7 Universities in India play a modified die-in-the-cup game that is played privately, and consecutively, 42 times. Subjects are asked to report the number of times, each of the numbers 1–6, on a six-sided die, turn up. Payoffs are increasing in the reported numbers except for the number 6 where the payoff is zero. The resulting temporal distribution, for each student, can then be compared against the statistical uniform distribution that is predicted to arise under truth-telling. This comparison can then reveal, unlike the static die-in-a-cup game, if each student lied or not. The students were then asked if they prefer government/public-sector jobs. Those who reported above-median scores on the die, were also 6.2% more likely to choose such jobs. A similar association is found between those who have less pro-social preferences, as revealed by their actions in a dictator game. There is no correlation between cognitive ability and cheating. However, the measure of cognitive ability used in the study differs from the measures used in the civil services exam in India.

The authors believe that these results speak to the screening and self-selection of individuals applying for government jobs. In a separate task, conducted with nurses in the public sector, in an identical game, the authors find that nurses who claimed above-median die scores are 10.7% more likely to engage in fraudulent absences from work. These results are potentially interesting, but the quantitative effects are weak. Although the study with nurses is useful in its own right, it is not clear how representative this particular subject pool is for those that apply for civil services exams in India.

4. External validity of lab evidence

The external validity of lab experiments has been an active area of research in behavioral economics. There have been concerns about the external validity of lab experiments (Levitt and List, 2007), but the emerging consensus is that lab evidence has a high degree of external validity that is possibly no different from the external validity of field experiments themselves (Fréchette and Schotter, 2015; Camerer, 2015; Dhami, 2016).

The subjects in Dai et al. (2017) are passengers using public transport in Lyon, France. They examine the correlation between fare evasion by passengers and their self-reported outcomes in the die-in-the-cup paradigm. Subjects were selected from the passengers who had just arrived at a tram/bus stop, and volunteered to participate in the experiment. The following three measures of dishonesty in the field were used. (1) At the end of the experiment, subjects were given the opportunity to exchange their ticket for a day pass, obviously a superior option. If they could not produce a ticket, they could have been fare-dodgers. (2) Subjects were also asked to self-report the number of times they evaded the fare for every 10 trips in the past; cheaters (or *self-fraudsters*) are classified as those who evade the fare at least once, while *non-fraudsters* are the rest. (3) A third measure was constructed by gathering data from those passengers who had just paid a fine for dodging the fare.

Each of these groups of people was then asked to participate in a die-in-the-cup lab experiment. In a slight departure from the standard experiment, the six faces of the die were given 3 colors, red, blue, and yellow, so there is a 1/3 probability that any of the three colors comes up in a random throw of the die. Subjects were asked to privately roll the die and self-report the outcome. The rewards were as follows: blue (0 Euros), yellow (3 Euros), and red (5 Euros). Clearly Econs should only report the color red.

There was widespread cheating behavior. When asked to produce a valid ticket to exchange for a day pass, 41.8% could not produce one. On the basis of self-reports, 54.92% travelled without a ticket once every 10 trips. Figure 4.1 shows the self-reports on the die throwing task for several categories of subjects, based on their behavior in the field. Thus, it depicts the relation between behaviors in the field and the lab. The categories are self-explanatory, based on our description above. For instance, the category non-ticket/Self-fraudster refers to those subjects who could not produce a ticket in exchange for the day pass, and who confessed to evading the fare at least once in the last 10 trips. The Figure also shows a horizontal line at 33.33%, which is the predicted statistical probability if everyone tells the truth.

The results are as follows.

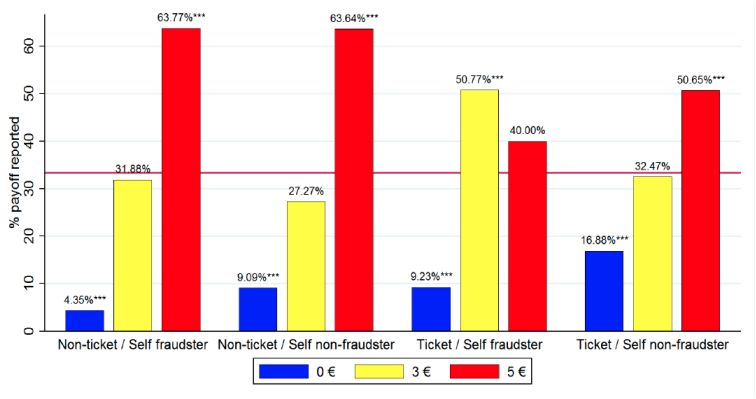


Figure 4.1: Results for the self-reports on the die throwing task in Dai et al. (2017), for 4 different categores of field subjects.

1. For all the 4 categories of people reported in Figure 4.1, the outcome with the highest payoff (in red) is over-reported, and the worst outcome (in blue) is underreported. Thus, in each category, the observed distribution is statistically different from the predicted uniform distribution.

2. Subjects who evade fare in the field are also more dishonest in the lab. Comparing the data for ticket holders and non-ticket holders, a p-test shows that the latter lie significantly more, i.e., overreport the best outcome (in red) and underreport the worst (in blue). Self-reported fraudsters underreport the worst outcome (in blue) significantly more than self-reported non-fraudsters, but they exhibit no difference in reporting the best outcome (in red).

3. The statistical distribution of reports for those who self-report never travelling without a ticket in the last 10 trips is significantly different, and more honest, relative to those who self-report travelling without a ticket at least thrice in the last 10 trips.

4. Those who have just been caught evading fare behave honestly in the lab experiment. The distribution of their reports is statistically indistinguishable from ticket holders. The experimental design is not rich enough to determine the reasons for this, nor speculate on how long-lasting these effects are. One may conjecture that some sort of *conscience-accounting* may be part of the explanation (see Gneezy et al., 2014, below).

5. When the die task is replaced by a contextualized lab public transport game that allows for fare evasion, then self-reported fraudsters in the field also behave more dishonestly than the rest.

Overall, these results suggest that lab behavior has a high degree of external validity.

The following two results indicate stability in preferences for lying over different subject pools.

Alm et al. (2015) compare the lab behavior of student and non-student populations in a tax evasion experiment in which subjects are informed about the probability of an

audit and the penalty rate. They find that although the mean compliance levels differ, the distributions of the compliance rates for the two groups are statistically identical. Further, the behavioral responses to changes in the compliance parameters are also identical for the two groups .

Armantier and Boly (2013) consider a corruption experiment (candidate offers a bribe to a grader to increase the grades) in three different settings–a lab in a developing country, a lab in a developed country, and in the field in a developing country. They find similar qualitative and quantitative results (for instance, to questions such as: what is the probability of bribes when the grader's wage increases?) when they compare the three different groups of subjects.

5. Gender differences in lying

Several studies find that men are more dishonest than women. The evidence comes from dishonesty tasks (Alm et al., 2009), tasks involving moral costs of lying (Fosgaard et al., 2013; Friesen and Gangadharan, 2012; Erat and Gneezy, 2012), lying in sender-receiver games (Dreber and Johannesson, 2008), self reports of coin tosses (Houser et al., 2012), return of excess change in restaurants (Azar et al., 2013), and fare dodging in a field experiment (Bucciol et al., 2013).

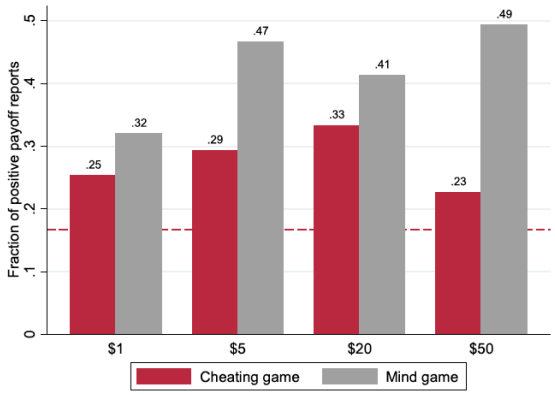
In an interesting experiment, Houser et al. (2016) explore some of the microfoundations of gender differences in lying behavior. Parents (88% mothers) were asked to toss two coins; each coin had a green side and a blue side. If both coins came up green, when tossed privately by each parent, then the parent was eligible for a reward. Since the outcome of the toss was not observed by the experimenter, individual lying was unobservable to the experimenter. However, one could compare the distribution of the claimed win rates relative to the objective probability of 25% of winning the prize under truthtelling. The authors chose a 2×2 design: (1) The reward was for either the parent ($10), or the child (a toy), and (2) the decision to lie was either made privately, or in front of the child, whose gender was recorded by the experimenter. The results were as follows:

1. Parents lied less in the presence of their child. When parents reported the outcome privately, the claimed win rate was 46%. However, in the presence of the child, this dropped down to 33%; the difference is significant at 10% ($p = 0.09$).

2. When the reward is for the parent (respectively, for the child), the claimed win rate was 36% (respectively, 43%); the difference is not significant ($p > 0.10$).

3. The greatest claim rates, 58%, occur when parents report privately, but the reward is for the child. This is statistically higher than the average claim rate of 33%, averaged across all treatments.

4. The claimed win rate in the presence of a daughter is 28%, close to the predicted rate of 25% under complete honesty. This is significantly lower than the claimed average win rate of 42% in the presence of sons ($p < 0.01$). This is perhaps the most interesting finding in the paper. It suggests that a potential explanation for lower



dishonesty among women may lie in the manner in which they are socialized when young, relative to men. The authors tie this to a result in Hays and Carver (2014), which shows that children who are exposed to dishonest behavior when young, are more likely to be dishonest when adults. However, this still begs the question of why people chose to behave differently with daughters, relative to sons, on moral issues.

6. Incentives and Lying

In sender-receiver games, the incentive to lie influences the extent of lying, suggesting that incentives may loosen one's morality (Gneezy, 2005; Sutter, 2009; Erat and Gneezy, 2012). In other methods, such as the matrix task and the die-in-a-cup method, cheating does not appear to respond to the extent of incentives. What accounts for this difference?

In Mazar et al. (2008), when subjects self-report the number of correctly solved matrices, at low incentive levels, \$0.10 and \$0.50 per correct matrix, there is a small level of dishonesty, relative to a baseline treatment. However, at higher incentive levels, \$2.50 and \$5.00 per solved matrix, there is no lying. The authors interpret these results as the outcome of a convex cost of lying, so that when lying for higher amounts, the marginal cost of lying to an individual increases. Using the die-in-a-cup method, Fischbacher and Föllmi-Heusi (2013) find that even when incentives are tripled, the extent of lying does not change. These general findings are supported in the meta-analysis of Abeler et al. (2016). However, when subjects are told that lying is legal, then the extent of lying responds to incentives (Gibson et al., 2013).

Kajackaitea and Gneezy (2017) try to reconcile these results by postulating that subjects who lie in the die-in-a-cup game may assign some residual probability of being found out. In order to test this idea, they propose a new game, the *mind game*, in which subjects first think of a number, then they roll a six-sided die. If they report that the number on the die is identical to the number they originally thought of, then they receive a prize that was varied in different treatments to reflect changing incentives (\$1, \$5, \$20, \$50).

The results on the fraction of claimed wins are shown in Figure 6 for each of the 4 incentive levels for the *mind game* and the usual die-in-a-cup method (which is referred to as the *cheating game*). The baseline level of 16.7%, which captures the statistically predicted proportion under truth-telling in the mind game (when the number on the die

is identical to the one thought of originally) is also shown in the diagram. The results are as follows:

1. There is greater lying in the mind game relative to the cheating game for all levels of incentives.

2. The percentage of participants who lie in the cheating game at the incentive levels \$1, \$5, \$20, \$50 is, respectively, 25, 29, 33, 23. There is no trend and, in fact, the lowest level of lying occurs when the incentives are the highest.

3. In the mind game, the percentage of participants who lie at the incentive levels \$1, \$5, \$20, \$50 is, respectively, 32, 47, 41, 49. The difference in cheating rates between the stake sizes \$5 and \$50 is statistically significant, although none of the other pairwise comparison of lying behavior has statistical significance. Lying, however, is significant (relative to the truthtelling benchmark of 16.7%) for all incentive levels. While men lie significantly more than women in the cheating game, there are no gender differences in lying in the mind game.

These results also appear to have significance for a different class of problems. Namely, that subjects might bring into the lab, norms and instincts from outside the lab. So, although the experimental instructions in the die-in-a-cup paradigm should assure subjects that there is no probability of being caught, this is not borne out by the extra cheating in the mind game. These results deserve to be replicated and studied further.

7. Some explanations for moral behavior

7.1. Maintenance of self-image

Mazar et al. (2008) proposed a theory of self-concept, or self-image maintenance; see also, Allport (1955) and Rosenberg (1979) for antecedents. In this theory, people have in mind some reference standard of behavior, say, relating to the desired degree of honesty. This reference standard, possibly context dependent, could conceivably be influenced by social norms for such behavior, or by one's own internal moral compass. When individuals take an action that falls below the reference standard, then they negatively update their self-image, which is aversive. Conversely, when actions exceed reference standard, individuals might positively update their self-image. These standards of behavior may be clearly categorizable on some ethical criteria, or they may fall into ambiguous categories, in which case, one may engage in self-serving justifications and rationalizations of having met the standards (Gino & Ariely, 2012; Shalvi et al. 2011).

Self-image may be malleable to the extent that one can be dishonest up to a limit, without having any adverse effect on one's self-image. But as this limit is exceeded, individuals negatively update their self image. This could explain, for instance, why people engage in partial lying instead of maximal lying, a common finding in the Fischbacher and Föllmi-Heusi type experiments. Splitting the benefits of a dishonest action with others may also produce a less negative update to the self-concept (Wiltermuth, 2011), as does the telling of White lies (Erat and Gneezy, 2012; Gino and Pierce, 2010; Gino et al., 2013).

7.2. Cost of lying

Individuals could incur a direct cost of lying, i.e., a direct moral cost that is subtracted from the utility they derive from an action (Ellingsen and Johannesson, 2004; Kartik, 2009). Or it could be that they suffer psychological costs such as guilt-aversion that inhibit lying (Charness and Dufwenberg, 2006); these costs arise from letting down the expectations of others that one is honest, i.e., they hinge on the second order beliefs of players. Abeler et al. (2016) find evidence for a direct preference for being honest and for being seen to be honest. One may also be subject to the trade-off between getting higher material payoffs through lying and reduced utility on account of moral transgressions and a deterioration in self-image; this has been termed as *ethical dissonance* (Barkan et al., 2012).

Using a sender-receiver game in Bangladesh, Leibbrandt et al. (2017) introduce the option to remain silent (rather than send either a true message or a false message). At high levels of stakes, worth about several months average wage, they find that this reduces the likelihood of sending a true message by 30%. However, there is no difference in the likelihood of sending a false message, thus, the option to remain silent is often exercised. These data are not consistent with the theory that people are intrinsically honest, but rather that there is a cost of lying that is balanced against the benefit of being virtuous. Another interesting feature of this dataset, as compared to the datasets from sender-receiver games in the Western world, is that only 54% of the receivers actually follow the sender's recommendations (compare this with 82% in Gneezy, 2005). Senders were significantly more optimistic and believed that 67% of the receivers would follow their recommendations; while 55% of the messages were true, receivers, on average, believed that 48% would be true, indicating that they probably took account of the option of remaining silent.

7.3. Moral balancing

Gneezy et al. (2014) provide another explanation for moral actions. Individuals might have self-imposed moral standards of behavior. If these are transgressed, then the individual might wish to engage in compensatory behavior, say, on account of *guilt*, to undo past transgressions (*moral balancing*). The authors give two nice examples of religious practices that tap into such a desire– the practice of *ashamot* (guilt) offerings as atonement for past transgressions in the Jewish faith, and *tariff penances* that date back to the medieval Catholic Church. Immediately following a transgression, in the hot state, one experiences a high level of guilt, but guilt depreciates over time as one enters a cold state. Individuals may also be forward-looking and take account of the subsequent depreciation of guilt, in the process adjusting the level of their current transgressions.

The authors consider a four period problem, so time $t = 1, 2, 3, 4$. A time t, the individual experiences time-dependent guilt, g_t, caused by potential transgressions. In contrast, the consumption allocations of the individual, x, and of others, y, materialize only in the last period, at time $t = 4$. The intertemporal utility, U, of the individual is given by

$$U = \sum\nolimits_{t=1}^{t=4} u(x, y, g_t),$$

where u is the instantaneous, and time-invariant, utility at time t, which satisfies the following conditions. Guilt is aversive ($u_g < 0$); marginal utility of consumption is reduced when one experiences more guilt ($u_{x,g} < 0$); and marginal disutility of guilt is reduced when the consumption of others increases ($u_{g,y} \geq 0$). The evolution of guilt takes the following form:

$$g_{t+1} = dg_t + a_t,$$

where $d \in (0, 1)$ captures the rate at which guilt decays, and a_t is a binary variable that takes a value 1 if a moral transgression is made and zero otherwise.

This simple model leads to the following set of testable predictions.

1. Since guilt depreciates over time, if an individual is given an opportunity to contribute to charity immediately after a moral transgression, he is more likely to give, relative to a delayed opportunity to give to a charity–this is the *conscience accounting hypothesis*.

2. If an individual knows that there will be an opportunity to donate to a charity after the moral transgression, this is likely to influence the extent of the moral transgression.

3. Prior to the transgression, if an individual is given the opportunity to choose between an immediate and delayed donation option, then awareness of the conscience accounting hypothesis may lead the individual to choose the delayed option. This might also separate individuals into two types depending on their moral preferences. Those who are worried about making decisions in a hot state might commit to choosing the delayed option. Others, who, say, have a high disutility from guilt, might commit to choosing the early option. It is also possible that subjects face a time consistency problem. Ex-post, those who have ex-ante chosen the early option, might regret and wish they had instead chosen the delayed option.

These predictions are confirmed by Gneezy et al. (2014) by adapting the sender-receiver framework of Gneezy (2005) described above; a dishonest message constitutes a moral transgression. In the *incentive choice treatment*, early and delayed options to contribute to charity were given to senders at the time they decide to send their message. In the *incentive and no incentive treatments*, the option to contribute to charity in an early/delayed manner was unexpectedly given once the senders had sent a message. In the *incentive reverse treatment*, the donation option was presented to subjects before they chose their messages and had no knowledge of the subsequent deception game; this feature differentiates this treatment from the incentive choice treatment. In the no incentive treatment, the sender does not have any incentive to lie, but in all the other incentive treatments, there is an incentive to lie. The results are as follows.

1. In the incentive treatment, the donation option arose unexpectedly once a sender had made the decision to lie. Of those who told the truth, 30% chose the donation option, and of those who lied, 73% chose the donation option; the differences are statistically significant.

Consider the incentive delay treatment in which subjects make a donation decision in a cold state after sufficient time has elapsed since they sent the original message (which could have been truthful or a lie). Subjects who sent a truthful message chose to make a donation in 33% of the cases, while those who lied chose to donate in 52% of the cases. Keeping payoffs fixed in the incentive treatments, subjects donated significantly less in the delay treatment as compared to the early donation treatment. This provides support for the conscience accounting hypothesis.

2. Consider the case when the donation decision was known before the sender had a chance to send a truthful or untruthful message. In this case, senders might be more willing to lie, in the knowledge that they can donate and pay for their transgression (*paying for sins hypothesis*). Here, 63% of the senders lied, a statistically higher percentage relative to the 48% who lied in the baseline treatment where no option to donate was given. Further, of those who lied in this treatment, 82% donate.
In the incentive choice treatment, when given a choice, 43% chose the option to donate early and 57% chose the delayed donation option. However, not all senders chose to honor their prior decision to donate. Comparing the senders who lied in these two treatments, 90% (early donation) and 31% (delayed donation) actually made the donation. Thus, donations are more likely to be made in the hot state.

Ploner and Regner (2017) ask subjects to first privately roll a six-sided die. Subjects who report an odd (even) number are entitled to play a dictator game with a higher (lower) endowment. The statistical probability of each report under truthtelling is 1/2, however, significantly more than 50% of the subjects report an odd number. Dictators who lied earlier (by claiming an odd number) transfer more money to the receiver relative to a baseline dictator game with an identical endowment where there is no possibility to cheat (the die is rolled in the presence of the experimenter). This suggests moral balancing.

Confessions may also be thought of as a form of moral balancing, in which one confesses to reduce guilt from an immoral act, or simply to morally cleanse oneself. However, evidence suggests that like partial lying in the Fischbacher and Föllmi-Heusi method, subjects engage in partial confessions, taking only *partial blame* for immoral acts (Pe'er et al., 2014).

7.4. Self-serving justifications

People might engage in self-serving justifications to weaken the apparent immorality of their actions, particularly when the moral benchmark might not be fully clear (Shalvi et al., 2011; Shalvi et al., 2012; Shalvi and Leiser, 2013). For instance, when people privately roll a die in the Fischbacher and Föllmi-Heusi method, they report lower numbers relative to a treatment in which they can roll the die thrice but are asked to report only the outcome for the first throw. The extra throws should be irrelevant for someone who wishes to tell the truth. However, for a potential liar, the situation is different. In contrast to lying about a high number on a single throw of the die, it might be considered more morally justifiable to report a higher number taken from the second or third throws of the die, even though one is asked to report the number only on the first throw. Similarly, when

individuals can justify immoral actions that benefits others, they are more likely to lie (Conrads et al., 2013).

7.5. Public personas, private personas, and morality

Like most other kinds of preferences identified in behavioral economics, morality is likely to be context dependent. Gintis (2017, Chapter 3) makes an important distinction between the *private persona* and the *public persona* of individuals that arise, respectively, in the *private sphere* and the *public spheres* of their actions. In the private sphere, individuals engage in private everyday transactions that may involve questions such as the following: Which consumer durables to buy? How to allocate the portfolio among alternative assets? How much to save? When to retire? The public sphere is defined as (Gintis, 2017, p. 47): "...the locus of activities that create, maintain, transform, interpret, enforce, and execute the rules of the game that define society itself." Examples include actions such as voting in elections, participating in a civil rights movements, and signing a petition for a social cause. The distinguishing feature of actions in the public sphere is that they are non-consequentialistic; they give rise to no private material payoffs, nor does any individual action, on its own, alter social outcomes. For instance, in signing a petition to ban fox hunting, one person's signature is unlikely to have an effect on the final outcome.

Individuals appear to behave 'as if' they put on different hats in the private and public spheres. In the private sphere, and under self-regarding preferences, individuals have private personas that are predicted to behave like the Econs in neoclassical economics. However, in the public sphere, individuals appear to have public personas, and derive direct utility from participating in actions in the public sphere. For instance, individuals might derive direct utility from voting in elections or from participating in social movements. However, such a preference is not absolute. Individuals could weigh the extra utility from these actions against the extra cost. So, the extra costs of voting, or participating in social actions may be high enough to dissuade some/many individuals from engaging in such actions.

In the light of this distinction, much of rational choice theory that is devoted to making sense of voting and participation in social actions is simply based on the incorrect assumption. Namely, that individuals take purely consequentialistic actions by engage their private personas. Gintis suggests that the appropriate equilibrium notion in the public sphere is a form of *social rationality*, as encapsulated in a Kantian equilibrium. In a symmetric n-player game, a Kantian equilibrium strategy is such that every player prefers it to all other strategies if "everyone who shares their preferences were to act according to the same rule." (Gintis, 2017, p. 51).

Dhami and al-Nowaihi (2010a,b) consider a theoretical model of behavioral political economy in which voters have Fehr-Schmidt other-regarding preferences (Fehr and Schmidt, 1997). They show that such voters behave in a self-regarding manner when choosing their individual labor supply, but behave in an other-regarding manner when choosing societal redistribution through voting. This is observationally equivalent to having a private persona in one sphere and a public persona in the other and as such provides some microfoundations to the idea. An analogy might help: A rich voter may send his

own children to a private school, but also vote for more public funds for state education. Dufwenberg et al., (2011) showed that this feature applies to a more general class of social preferences (Dhami, 2016, Sections 6.5, 6.6).

8. Exploring the richness of human morality

8.1. Are there two types of liars or several?

Hurkens and Kartik (2009) reconsider the results of Gneezy (2005) and show that they are unable to reject the hypothesis that the data came from a population of players of two types. Type-I never lies whatever the cost (ethical type) and Type-II always lies (unethical type). However, if the types are fixed, this evidence begs the question of why the extent of lying responds to the incentives to lie in so many diverse contexts.

Gibson et al. (2013) conduct a decision-theoretic lab experiment in which the confounds of strategic considerations and other-regarding preferences are eliminated. In the truth-telling task, subjects, in their role as the CEO of a company, had to decide on two possible earnings announcements that affected their own payoffs; a higher announcement is a lie, but it also increases their payoff. In several different treatments, the payoff differences between truth-telling and lying are different. For instance, in one treatment, the choice was between announcing 31 cents/share and 35 cents/share; the higher announcement led to 5 times higher earnings for the CEO. The subjects are told that the lie is within the bounds of what could be defended on accounting grounds. Thus, one important confounding influence in this study is that subjects are made to believe that lying is legal.

If the subjects were of only two types, ethical and unethical, then we should not get any variation in the levels of truthfulness as the incentive to lie varies. The ethical types should never lie and the unethical types should always lie. In contrast, the extent of lying was sensitive to the incentive to lie. The authors find that the most significant factor in the decision to lie is the intrinsic cost of lying, which ties-in with the results of Gneezy (2005). Hence, people seem to have underlying preferences over how much they are willing to lie as incentives for lying vary.

Gneezy et al. (2013) classify subjects into 8 types, depending on the extent of their lying. The main types are as follows. Some subjects are always honest, irrespective of the incentives. Others always maximize their monetary benefits (similar to the amoral, and self-regarding preferences in neoclassical economics). In a sender-receiver game, over all periods, the authors find that the respective percentages of these two groups among the senders of the message are 33% and 28%. Another group responds to the incentives to lie–lying more when the incentives to do so are high. Interestingly, as subjects gain experience of lying, they lie more. This ties in with the *slippery slope of dishonesty* (see Section 12). An alternative explanation may also be given in terms of the *depletion of self-control and willpower* upon repeated truth-telling in the face of a temptation to lie (Mead et al. 2009; Gino et al., 2011). Future research may try to distinguish between these two alternative explanations.

8.2. Lying in gain and loss frames

In the spirit of Kahneman and Tversky's (1979) prospect theory, morality might also be influenced by whether, relative to some *reference point*, one is in a *gain frame* or in a *loss frame*. This is because, due to *loss-aversion*, losses typically bite, on average, about 2.25 times equivalent gains–a robust finding in humans and close primate relatives such as capuchin monkey (Dhami, 2016, Chapters 3, 20, 21). For instance, school teachers work harder to enable students to achieve higher grades when an up-front bonus is paid to them that could be clawed back if grades fall (loss aversion), as compared to a bonus that is paid in the end, once grades materialize (Fryer, 2012). Insofar as enabling students to achieve their potential is considered to be a moral duty of teachers, this result also speaks to the differing effects of morality in the loss and gain frames.

More direct evidence on morality in gain and loss frames is provided by Schindler and Pfattheicher (2017). They use a variant of the Fischbacher-Föllmi-Heusi method and ask subjects to privately roll a die 75 times and report the number of occurrences of 4. Since the probability of a 4 in each throw is 1/6 and the throws are independent, the statistical prediction under truth-telling is $75\,(1/6) = 12.5$. In the gain frame, subjects are told that they will gain 10 cents for each reported 4. In the loss frame, subjects are initially endowed with 7.5 euros and told that they will lose 10 cents for every report of a number that is not 4. This is identical to the opportunity cost of not reporting a 4 in the gain frame. On average, subjects in the loss frame reported significantly higher 4s as compared to truth-telling ($p = 0.031$), which indicates significant dishonesty. In contrast, in the gain frame, no statistically significant dishonesty was found.

Grolleau et al. (2016) give subjects a matrix solving task in a 2×2 design (gain vs loss frame and monitored vs unmonitored reporting). In the gain frame, subjects are given a payment for the number of correct solutions. In the loss frame, subjects are initially given the maximum possible payment, which corresponds to correctly solving all the matrices. Then, based on their actual/reported performance in the matrix task, payment is clawed back from them for the unsolved matrices.

In the monitored treatment, where no cheating is possible, there is no significant difference in performance between the gain and the loss frames. Thus, the frames do not produce any innate differences in the motivation to solve extra matrices. However, in the unmonitored treatment, when cheating to the full extent is possible, without any detection, there is a marked difference in cheating in the two frames. Relative to the monitored frame, in the unmonitored frame the percentage of solved matrices increases by 43% in the gains frame and by 296% in the loss frame. Both differences are significant, but the differential effect under the loss frame is several orders of magnitude higher.

Several other papers also report an increase in cheating when subjects are in the loss frame relative to the gain frame. The opportunity to convert a loss into a gain induces greater cheating (Shalvi, 2012). Goals might serve as a reference point and falling below the goals presumably puts people in a loss frame (Dhami, 2016; Part 1, Section 3.7). It is found that there is more unethical behavior when subjects fall below their goals (Schweitzer et al. 2004).

Garbarino et al., (2017) consider the relation between loss aversion and the probability

of receiving various outcomes. Suppose that there are only two outcomes, $x_1 < x_2$, received with respective probabilities p, $1-p$ ($p \in (0,1)$). The decision maker has the option to report either of the two outcomes in a truthtelling task. Assume that the reference point, in the sense of prospect theory, is the average outcome, $\overline{x} = px_1 + (1-p)x_2$, where $x_1 < \overline{x} < x_2$. Suppose now that p increases, then the reference point $\overline{x}$ increases too. Let the individual be loss averse. The increase in the reference point implies that reporting the lower outcome in the truthtelling task will reduce the individual's utility proportionately more (on account of loss-aversion). Thus, the incentive to lie is sensitive to the probability distribution of outcomes. The authors confirm this prediction, while using data from 81 studies. They also suggest new econometric techniques to uncover the distribution of lying individuals in the sample.

8.3. Delegation and third party punishment

Third party punishment is an enduring feature of human behavior. Humans appear to be hardwired with this feature, and it is an essential component in the maintenance of human morality and social norms (Gintis, 2009, 2017; Dhami, 2016, Part 2). Fehr and Gächter (2000) showed that in public goods games, contributors engage in costly third party punishment of non-contributors (pro-social punishment). This was replicated in a large number of experiments conducted in the West. However, in data gathered in the rest of the world it was shown that such punishment can also take an anti-social form (non-contributors punish contributors as revenge for past punishments). It turns out that norms of civic cooperation (which encompass attitudes to tax evasion and abuse of the welfare state) and the rule of law (which reflects peoples' trust in law enforcement institutions) are positively correlated with pro-social punishment (Herrmann et al., 2008).

Bartling et al., (2014) identify the importance of third party punishment in the context of moral behavior. In their experiments, dictators could, through their actions that always benefit themselves, have one of two effects on the receivers– a beneficial or a harmful effect. Dictators could, costlessly and voluntarily, choose to be informed, or stay ignorant of the effect on the receivers. A third party observes the choice of the dictator and decides whether or not to punish the dictator. Third party Econs would never engage in such punishment because bygones are bygones (Dhami, 2016; Section 6.2, p.52).

Suppose that the third party observes a harmful effect on the receiver. Then, conditional on the dictator having chosen to remain ignorant, the evidence shows that s/he is punished less. Thus, ignorance helps to reduce the blame for an unfair outcome. So why don't people choose to stay wilfully ignorant all the time? The answer is provided by comparing the punishments that the third parties impose on the dictators when the dictators chose to be ignorant, relative to being informed. Punishments are significantly higher in the former case. Thus, on net, it might not help dictators to stay ignorant. Since the dictator's decision reveals the dictator's intentions, a model of intentions based reciprocity is probably the ideal vehicle to purse these ideas further.

Bartling and Fischbacher (2012) ask if blame for unpleasant outcomes can be reduced by delegating decisions in a dictator game to a third party. They consider a four-player variant of a standard two-player dictator game. Of the four players, labeled as A, B, C,

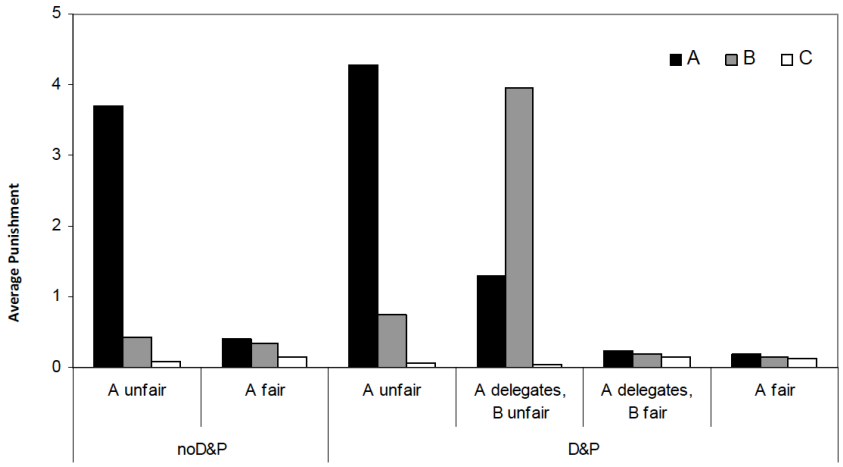


FIGURE 1. PUNISHMENT PATTERN IN TREATMENTS *noD&P* AND *D&P*

C, player A is the dictator who can choose to delegate the division of a fixed pie of size 20, player B is the potential delegee, and players C, C are the two receivers who may or may not be able to punish unfair allocations depending on the treatment. Only one of two possible allocations can be offered to the receivers– a fair allocation that gives 5 units to all four players and an unfair allocation that gives 9 units each to A and B and 1 unit each to the two receivers.

There are pros and cons of delegation for player A. On the one hand, it may allow player to deflect blame for an unfair allocation, but, on the other, the delegee may choose a fair allocation and reduce player As payoff.

The first set of results is shown in Figure 8.3. There is a set of (1) two histograms for the treatment noD&P (no delegation allowed but one of the player Cs can punish) and differentiated by whether A chooses a fair or unfair option, and (2) four histograms for the treatment D&P (delegation is allowed and one of the player Cs can punish) differentiated by fair and unfair offers and by delegation or no delegation. If A chooses a fair outcome, then there is no significant difference in punishment among the two treatments (compare the second and sixth histograms). However, it is not clear why there should be any punishment at all in this case.

Conditional on an unfair outcome, the main target of the punishment is the individual who decides the unfair outcome (A in treatment noD&P and if no delegation occurs in D&P; and B if delegation occurs in D&P). On comparing the fair and unfair outcomes cases, the punishment levels are significantly higher under unfair outcomes ($p < 0.01$ in a two-sided Wilcoxon signed rank test). It appears that blame can be successfully shifted. Conditional on an unfair outcome, the average punishment meted out to A is 4.27 points in the absence of delegation and 1.13 points in the presence of delegation–a statistically significant difference.

In order to explore further aspects of morality, the authors design variants of the D&P treatment. In the treatment random, A can delegate the decision to a random device that chooses the unfair outcome with a 40% chance and the fair outcome with a 60% chance. In the treatment asymmetric, A can choose either the fair allocation (but not the unfair

allocation) or delegate to B.

The results are as follows. Conditional on an unfair outcome, A is punished less if he chose the unfair outcome as compared to delegating to the random device. Thus, even delegation to a random device appears to shift some of the blame for an unfair outcome away from A. This result stands in contrast to the well-known result of Blount (1995) who showed that in ultimatum games, low offers are more likely to be accepted if chosen by a random device. In the asymmetric treatment where A cannot choose an unfair outcome, and conditional on an unfair outcome (which can only be chosen by the delegee, B), player A is punished relative more as compared to delegation to B under treatment D&P. Perhaps this allows player Cs to better infer the unkind intentions of player A. Finally, over repeated interaction, player As learn that player Bs choose the unfair outcome sufficiently often that, in terms of expected payoffs, it pays to delegate the decision to B. In a nutshell, the moral aspects of delegation help explain why delegation occurs.

Erat (2013) also finds that people prefer to delegate lying to others. Further, when incentives to lie are varied, in terms of the harm that the lie causes to others, people delegate more as the harm to others increases. When the harm caused to others increases, women are more likely to delegate than men.

Whistle-blowers in organizations, a form of third-party punishment, often put themselves at substantial risk to report unethical behavior. Why are people willing to do so, and how do others perceive their behavior? Reuben and Stephenson (2013) investigate these issues in a lab experiment where subjects play a repeated whistleblowing game. In each round, students in a group observe a number that they can potentially overstate to receive a higher payoff. Group members observe the actual and reported numbers of all other group members, and then choose whether they would like to report cheating (whistleblowing). The whistleblower receives no benefit from reporting a liar, but the liar receives a monetary sanction that is increasing in the size of the lie. Since there is no private benefit from the reporting of dishonest activity, if any reports occur, they must be because lying, per se, is considered to be undesirable. Econs will not engage in the whistleblowing activity.

In order to investigate how others perceive whistleblowing, every three rounds, some of the groups are reshuffled by removing some members. The removed members may rejoin other groups, provided that they are accepted into a new group with a unanimous vote after their history of lying, whistleblowing, and earnings in previous rounds is revealed to new group members.

The results are as follows. There is sufficient whistleblowing activity, such that, in the average group, lying does not pay (recall reported liars pay penalties). The presence of whistle-blowers suggests that individuals have a preference to engage in moral activities, even when there are no private material gains. However, interestingly, whistle-blowers are less likely to be inducted into new groups, relative to liars. The reason could be that people do not like snitching behavior. This process leads to the formation of several highly dishonest groups in which there is little whistleblowing. In such groups, dishonesty pays, in the sense, that liars receive higher payoffs than honest subjects.

8.4. History-dependent lying

One may cheat more in a given situation if one feels unfairly treated in the past. In order to test this intuition, Houser et al. (2012) first play a dictator game with their subjects who are randomly assigned to be either dictators or receivers. Having played this game, subjects are then given an ethical choice. They are asked to privately toss a coin; a report of heads earns them 1 euro and a report of a tails earns 3 euros. There is significant lying; 74.5% report tails, while under truthtelling 50% should report tails. Using a mixture model with two types, those who lie and those who don't, the authors estimate that the cheating rate among receivers is higher than that among dictators (53% versus 45%). Furthermore, this difference is driven almost entirely by those receivers who received nothing in the dictator game. There is no noticeable difference in cheating between dictators and receivers who received a positive amount, despite the dictator's payoffs being greater.

8.5. Lying in groups

Do people lie more in groups or when they make individual decisions? What group features may induce more or less honesty? One cannot provide an answer to these questions by purely deductive reasoning because the arguments could go either way. (1) Groups may be able to use a more sophisticated analysis relative to individuals and ensure a better understanding of the underlying game (Kocher et al., 2006; Sutter, 2009). (2) Individuals might be able to disguise their lying in groups. (3) Individuals might also lie more in groups, on account of social preferences (Gino et al. 2013, Wiltermuth, 2011). (4) Concerns for one's social image might reduce lying in groups (Bénabou, 2013; Bénabou and Tirole, 2006). (5) Group interaction may reveal social norms about honesty, which could either increase or reduce honesty, depending on what one observes and learns.

This important class of questions is addressed by Kocher et al. (2017). In their novel use of the Fischbacher and Föllmi-Heusi method, subjects, in the *individual treatment*, observe the throw of a die on a computer screen and self-report the observed number; payoffs equal the number reported, except for the number 6, which results in a zero payoff. The experimenter also observes the outcome of the die on the computer screen. Hence, individual cheating behavior can be identified. Thus, relative to the standard Fischbacher and Föllmi-Heusi setup, this may create greater uncertainty on the part of subjects that their lying could be observed by others.

Once subjects have participated in the individual treatment, they then participate in either of two group cheating tasks. In the group tasks, subjects observe the throw of a die on the computer, but then have the opportunity to chat and exchange free-form messages before they individually submit their reported numbers. In the treatment GroupPC, payoffs are common, in the sense that if all subjects report the same number, then their individual payoffs equal the number (except for a payoff of zero for the number 6), otherwise all get a zero payoff. This treatment activates group concerns and social preferences. In contrast, in the treatment GroupNoPC, the payoff of each player depends only on the number reported by the player, so social concerns are absent.

Comparing the results across the three treatments, the main results are as follows.

1. There is significantly more lying in GroupPC (89.7%) and in GroupNoPC (86.3%)

as compared to lying in the individual treatment that precedes it (61.5%). The average percentages are significantly different between Individual and GroupPC, and Individual and Group NoPC. Thus, groups lie significantly more than individuals. There is no statistical difference in lying between the two group treatments, suggesting that social preferences are not important in this context. The choices made in the Individual treatment (honest or dishonest) have no bearing on the choices made in the group treatments. Once communication is allowed in the two group treatments, the cheating rate increases significantly.

2. There is a high degree of coordination among group members in both treatments. Coordination is in everyone's best interest in GroupPC, and all subjects are found to coordinate. There is also surprisingly high coordination in GroupNoPC, where coordination has no payoff relevance. Here 33 out of 39 groups coordinate after the group chat.

3. So what causes increased lying in groups? First, group chat increases the beliefs of players about the dishonesty of others, relative to the Individual treatment. Second, communication during the group chat plays a key role in the decision to lie. Research assistants are used to categorize free-form chat arguments into those that reflect honesty and dishonesty; when there is a difference of opinions among the research assistants, the median value among the research assistants is taken. Arguments for dishonesty are made far more frequently (in 51% of the groups) as compared to arguments for honesty (in only 24% of the groups).
   Examining individual arguments, it turns out that 43.4% of the arguments favor dishonesty while 15.6% favor honesty. The number of arguments for dishonesty are indistinguishable among the two groups. Thus, payoff commonality does not appear to be a factor in dishonest behavior in groups. Finally, arguments for honesty significantly reduce lying in groups.

In conjunction, these results suggest that there is a shift in the perception of individuals in groups about the honesty norms in the rest of the population that drives the results.

Balafoutas et al. (2017) conducted a real effort experiment with professional German internal auditors who are members of the German Institute for Internal Audit. Subjects were given a set of 30 calculations and told that 10 were incorrect. The task was to identify the number of incorrect calculations in 3 minutes, without using a calculator; 1 point was given for a correct identification and 0.5 points deducted for each incorrect identification.

In each of three different treatments, subjects received an identical show-up fee. However, the treatments differed in the incentives to lie. Under *individual incentives*, subjects received a piece rate of 2 euros for each point. Under *competitive incentives*, random groups of 2 individuals were formed and, in each group, the individual who gives more correct answers gets 4 euros for each point, while the partner gets nothing; the identity of the partner is kept anonymous. Under *team incentives*, each player gets 1 euro for a correct answer given by any of the two players in the team.

The number of correct answers was determined in two possible ways, in this 3×2 design (3 treatments and 2 different methods of evaluation). Under *objective evaluation*,

each player makes a report about the other's performance, but such reports have no payoff consequence and the experimenter directly checks the number of correct answers. Under *peer evaluation*, the payoff of a player is determined by the report of the partner. The reports under peer evaluation can be honest reports, or they can be lies, depending on the incentive structure (e.g., competitive or team incentives). The results are as follows:

1. Under *individual incentives* and *peer evaluation*, the actual performance (6.07 points) and the report on the performance made by peers (5.96) is statistically indistinguishable. Thus, there is no misreporting in the absence of monetary incentives.

2. Under *objective evaluation* and *team incentives*, players underreport the performance of the team member, despite their report having no bearing on the payoffs. The authors conjecture that perhaps this enhances the *self-image* and *status perception* of individuals. This result also nicely dovetails with Carpenter et al. (2010) and Charness et al. (2014) who find evidence that in the absence of monetary incentives, people who compete with each other may sabotage each other.

3. Under *peer evaluation* and *competitive incentives*, there is underreporting of the number of correct answers achieved by the partner. On average, reported points are 22% lower than actual points (on average, actual points are 5.87, reported points are 4.57, $p = 0.04$ in a Wilcoxon signed ranks-tests). The opposite, i.e., overreporting, occurs under *team incentives*; reports inflate the number of correct points by 16%.

4. Dishonest behavior is driven by a minority of the subjects, while most subjects are honest. The share of truthful reports ranges from 70% (under peer reporting and team incentives) to 86% (under peer reporting and individual incentives). However, the behavior of the small minority does lead to differences in the averages across treatments (as in points 2 and 3 above). The share of dishonest subjects is higher under competitive incentives and team incentives relative to individual incentives. Faravelli et al. (2015) also find that under competitive incentives, self-reported performance increases, as does the proportion of individuals who self-select themselves for such tasks.

8.6. Moral suasion and morality

People may be induced to act morally through simple *moral suasion*. Indeed, an older literature in the economics of banking used to stress the positive role of moral suasion in banking, whereby the Federal Reserve in the US used 'persuasion' (e.g., closed door meetings with bank directors and appealing to the public-spiritedness of actions) to informally regulate private banks as compared to formal and binding regulation (Breton and Wintrobe, 1978). Moral suasion is not predicted to have any effect on Econs, unless it reveals some relevant information, or if there are reputational issues at stake.

In two natural field experiments, Hallsworth et al. (2017) randomized 5 different messages across 100,000 taxpayers who had declared their incomes, but had not paid their taxes yet. Thus, the study is not about the taxpayers' decision to pay or not pay taxes, but rather, the timeliness of their payments. In the experiment, a control group received a

standard letter, but no moral persuasion was involved. In the letters where moral persuasion was involved, 3 of the 5 messages were norm-based (see examples below), while the remaining two were public service messages (e.g., "taxes fund public services"). Relative to the control treatment, the 5 messages resulted in an increase in the likelihood of an earlier repayment of taxes. The most successful of these messages produced a treatment effect of 5.1% over the control treatment.

In a second experiment, the authors distinguish between *descriptive norms* and *injunctive norms* (respectively, *empirical* and *normative* expectations in the terminology of Bicchieri, 2006). Descriptive norms are designed to tell subjects about what others do (e.g., "most other taxpayers pay with minimal delay") and injunctive norms tell subjects about what others think should be done (e.g., "most people believe that taxpayers should not delay payments beyond a month"). It is found that descriptive norms are relatively more effective in persuading taxpayers to pay early. Thus, moral suasion is effective, and we now know more about the form of moral suasion that is more effective.

8.7. Morality and social identity

There has been an explosion of research on social identity in economics; for a survey, see Dhami (2016, Ch. 7). People identify themselves with social categories, each potentially representing a distinct social identity; each social category may have its own norms and ideal behaviors. Once individuals associate with a social identity, then one observes favourable behavior towards ingroup members (other members of the same social category) and unfavorable behavior towards outgroup members or non-members (Tajfel et al., 1971; Tajfel and Turner, 1986; Akerlof and Kranton, 2005). For instance, what may constitute immoral and unacceptable behavior towards ingroup members might be perfectly acceptable when directed towards outgroup members. Killing of an ingroup member would, in most groups, be considered immoral, yet in times of wars, perhaps aided by propaganda, killing of outgroup members is considered valiant and praiseworthy.

Humans appear so hardwired to respond to ingroup/outgroup distinctions in behavior that when they are primed for even minimal group identities, such as red and blue groups, they favour ingroup members relative to outgroup members. The presence and persistence of stereotypes and of discrimination towards other groups can be explained along these lines. For instance, Bertrand and Mullainathan (2004) respond to help-wanted ads in Boston and Chicago newspapers, and for identical resumes, randomize among African-American and White sounding names. The later receive 50% more callbacks. In the trust game experiments of Eckel and Petrie (2011), subjects can view a photograph of the other player, at a price. Among those who chose to view the photograph, white trustors discriminate favorably towards white, relative to black trustees. Black trustors do not discriminate, perhaps because many of their role models, such as school teachers and doctors, are white. On the whole, there appears to be an information value in a face that may be explained in terms of social identity.

Human morality also appears tied to the *social or professional identity* that one assumes when making unethical decisions. This is nicely illustrated in the work of Cohn et al. (2014). They divide 200 bank employees into a control group and an experimental group.

Subjects in both groups privately toss a coin and, based on the privately reported outcome (heads or tails), they can increase their income by up to $200. The two groups are primed differently for their identity. The control group was primed for it's non-occupational identity by asking questions such as: What is your favorite activity in your leisure time? The experimental group was primed for its professional or occupational identity by asking questions such as: What bank do you work at? How long have you been working in the banking sector?

The results are as follows. In the control group, bank employees were honest. Compared to the statistical benchmark of 50% under full-honesty, 51.6% claimed a successful coin flip; the two percentages are not statistically different. However, subjects in the experimental group are significantly more dishonest as compared to those in the control group; 58.2% reported a successful coin flip, which is significantly different from the predicted 50% under truthtelling, and from 51.6% in the control group. In order to check if the increased dishonesty was specific to the banking sector, the authors then repeated the control and experimental conditions with 133 employees of other industries. For the non-banking employees, there was no difference in the honesty levels between the control and experimental conditions. The authors conclude that the results, in their sample, are driven by the existing banking culture which appears not to be fully honest. This inference is subject to the usual caveat in such field experiments that the employees were not a random sample of all possible bank employees.

Conducting experiments on prison inmates in a Swiss prison, Cohn et al. (2015) postulate that 'deviant people' have two identities–a criminal identity and a moral identity. Violations of rules imposes no costs on the criminal identity, but imposes costs on the moral identity. Hence, they argue, if criminals could be primed for their criminal identity (e.g., by asking: What are you convicted of?), then, relative to priming for a non-criminal identity (e.g., by asking: How many hours do you watch TV?), the psychological costs of rule violations are likely to be lower. Thus, a criminal identity is predicted to induce more rule-breaking behavior. Subjects were asked to flip 10 coins, privately, and they could keep any coins that they reported as heads. Since individual cheating could not be observed, the authors compare the reports to the statistical prediction of 50% heads under truth-telling.

Half the subjects were primed for their criminal identity, and the other half for a non-criminal identity (control condition). In the control condition, they find that 60% reported heads; thus, $2(60-50) = 20$ percent of the reports were lies. When the subjects are primed for criminal identity, subjects reported 66% heads, so $2(66-50) = 32$ percent lied, which is significantly different from the control condition. Further, the authors find a positive correlation between lying in the coin toss task and lack of compliance of prisoner's with prison regulations, e.g., aggressive behavior towards others, and use of illegal drugs.

8.8. Morality and anonymity

Moral actions may be underpinned by emotions such as 'shame' that arise from others' observing our actions. In dictator games, Haley and Fessler (2005) show that when dictators are shown pictures of eyes in the same room while making decisions, they make more generous decisions. Perhaps the pictures induce a feeling of being watched, hence,

triggering emotions such as shame if dictators make low offers. In a different context, these results are supported by the empirical findings of Bateson et al. (2006).

Some people may cross over to the other side of a road when they see a beggar, to avoid the guilt that they would feel if they did not give something to the beggar.[7] Dana et al. (2006) and Dana et al. (2007) test a similar idea in the lab by giving dictators the opportunity to either remain anonymous or not, and to remain ignorant or not, in a dictator game.

When dictators are given an initial endowment of \$10 and also given the opportunity to exit the experiment with \$9, without the receivers ever finding out, 28% choose the exit option. However, the exiting dictators could have kept \$10 by playing the game and offering nothing to the receivers (Dana et al., 2006). In another experiment, dictators are given a choice between being aware and being ignorant of the payoffs of the receiver. A majority (56%) of the dictators chose to be ignorant when there is a likelihood that the payoffs could reveal a lower payoff to the receiver in the state where the dictator has a higher payoff (Dana et al., 2007). The authors term this as *moral wiggle room*, which is exercised by many participants. Thus, many people might be termed as *reluctant altruists*.

Despite its widespread usage, the dictator game is a very special game in which only one party makes a decision of any significance. The most direct real world analogue of the dictator game, contributions to charity, is also suspect because modern charities use active strategies to solicit contributions. Even helping out a beggar on the street might be a poor analogue of a dictator game because the condition of a beggar and his pleas may elicit empathy and guilt, which are not a part of the anonymous experimental dictator game. This suggests interpreting the results from dictator game with caution. Indeed, many of the results from dictator game experiments do not survive in the presence of strategic interaction (Fehr and Schmidt, 2007; Dhami, 2016). In trust games, the eyes cue (as in the Haley-Fessler experiments reported above) has no effect on the degree of prosociality of offers (Fehr and Schneider, 2010). In the trust and the moonlighting games, the presence of moral wiggle room does not reduce reciprocity (van der Weele et al., 2014).

Gneezy et al. (2017) found that people are unwilling to share their, possibly negative, views on the attractiveness of other people, even if shading their views comes at a personal material cost. When asked to share their views under anonymity, subjects are relatively truthful. The authors conclude that people do not wish to be messengers of bad news.

9. Morality and beliefs: Psychological game theory

Given a set of players, N, whose pure strategy profiles are given by the set S, classical game theory is mainly interested in the material utility of player i, $u_i : S \to R$.[8] Classical game theory has an important role for beliefs, e.g., beliefs are updated using Bayes' Law, whenever possible. However, the classical framework is not well suited to considering the

[7] On the other hand, an important explanation for why people give to charities is that they derive a *warm glow* from the act of giving (Andreoni, 1990). There has been support for this idea in some experiments and in neuroeconomic studies (Harbaugh et al., 2007). However, there is debate about the relative importance of 'pure altruism' and 'warm glow' in charitable giving.

[8] This discussion can be extended to mixed strategies.

role of a range of emotions, such as guilt-aversion, surprise-seeking, reciprocity, malice, anger, and shame-aversion, that underpin human morality and play a critical role in the development and upkeep of social and moral norms (Elster, 1998, 2011; Bicchieri, 2006).

In recent years, rapid progress has been made in psychological game theory (PGT), which allows beliefs to directly enter into the utility function of players. Let B be the hierarchy of beliefs of all orders for all players. Then, under PGT, the utility function of player i is given by $u_i : S \times B \to R$. This is not simply a matter of augmenting material payoffs with beliefs of various orders and then applying the classical machinery in game theory. This is because beliefs themselves may be endogenous, hence, an entirely new framework, *psychological game theory*, is needed.[9] The following example illustrates how the feelings of *surprise* and *guilt* may directly impart disutility.

Example 1 *: John frequently visits cities A and B, and he typically uses a taxi to get around. In city A, tipping a taxi driver is considered insulting, while in city B it is the norm to tip a publicly known percentage of the fare. Suppose that it is common knowledge that if taxi drivers do not receive a tip, they quietly drive away. In city A, John gives no tip, and feels no remorse from not giving it. However, in city B, the taxi driver expects John to give him a tip (taxi driver's first order belief) and John believes that the taxi driver expects a tip from him (John's second order belief). Based on his second order belief, John cannot bear the guilt of letting the taxi driver down by not paying the tip. Thus, he tips every time he takes a taxi in city B. Clearly, John's utility appears to be directly influenced by his second order beliefs.*

In Example 1, in city B, if John believes that the taxi driver has been particularly courteous and helpful, then he might tip him extra, on account of *reciprocity*. Following the pioneering work of Geanakoplos et al. (1989), Rabin (1993) showed how reciprocity could be formally modelled in simultaneous move games. This work was extended to sequential games by Dufwenberg and Kirchsteiger (2004) and then to a more general class of models by Battigalli and Dufwenberg (2009).

Battigalli and Dufwenberg (2007) proposed a formal approach to modelling guilt. They distinguish between two different emotions associated with guilt.

(1) *Simple guilt* arises from falling short of the perceived expectations of other players. For instance, if in city B in Example 1, John believes that the taxi driver expects a 15% tip, yet pays only a 10% tip, then he may suffer from simple guilt, which directly reduces his utility.

(2) *Guilt from blame* arises when one cares for the attribution of intentions behind psychological feelings such as guilt–aversion/surprise–seeking. In Example 1, suppose some taxi drivers who fail to receive a tip behave in an awkward and insulting manner. On observing a tip, the taxi driver must infer if John gave the tip purely on account of moral reasons (say, guilt-aversion) or because he preferred not to have an unpleasant argument with the taxi driver. Since guilt-aversion itself relies on second order beliefs, the taxi driver needs to form his third order beliefs about John's second order beliefs in order to form this inference. In turn, John may derive direct disutility if he believes that his tip was believed

[9] For a treatment of psychological game theory and more examples, see Section 13.5 in Dhami (2016).

by the taxi driver to be unintentional, in the sense that it was given to avoid a potential argument, rather than for moral reasons. However, since John does not observe the taxi driver's third order beliefs, he must form fourth order beliefs about the taxi driver's third order beliefs in order to form this inference.

Typically, models in PGT restrict themselves to analyzing beliefs upto order 4 because it does not seem compelling that most people have the cognitive ability to form beliefs of higher orders. In contrast, classical game theory makes the empirically rejected assumption that players can form beliefs upto any order; that there is common knowledge in the form of an infinite regress of beliefs; and that beliefs and actions are consistent with each other (Dhami, 2016, Part 4).

The surprise–seeking motive was formally identified by Khalmetski et al. (2015) in dictator game experiments. They also provide a theoretical framework in which surprise–seeking may be analyzed. The surprise–seeking motive arises from exceeding the expectations of others, as perceived by a player through his/her second order beliefs. For instance, in Example 1, in city B, John may believe that the taxi driver expects a tip that is 10% of the fare, yet he may derive extra utility by offering instead a 15% tip (surprise–seeking motive) that puts a smile on the taxi driver's face. One may extend these beliefs to higher orders by factoring in the intentionality of the surprise–seeking motive.

Empirical studies based on eliciting the beliefs of players by a self-reporting method (or *the direct elicitation method*) find strong support for the simple guilt–aversion motive in trust games and public goods games. Operationally, guilt-aversion is confirmed by a significant correlation coefficient between one's actions and one's second order beliefs (i.e., beliefs about the other players' first order beliefs).[10]

Ellingsen et al. (2010) question the validity of the self-reporting method. They argue that self-reported second order beliefs of players, i.e., beliefs about the first order beliefs of others, are subject to the *false consensus effect*, which is an example of *evidential reasoning* (Ross et al., 1977; al-Nowaihi and Dhami, 2015). They proposed instead, the *induced beliefs method*, to elicit beliefs. In the first stage, they directly ask players for their first order beliefs. These beliefs are then revealed to the other player before they make their decision. Players are given no information about how their beliefs will be used, so it is hoped that beliefs are not misstated to gain a strategic advantage. Thus, the second order beliefs of players (beliefs about the first order beliefs of others) are as accurate as possible. It is 'as if' players can peep into the minds of other players to accurately gauge their beliefs.[11] Using this method, they find that the correlation between second order beliefs and actions is not statistically different from zero, i.e., guilt-aversion is absent and is confounded by the false consensus effect.

Khalmetski et al. (2015) showed, in dictator game experiments, that the Ellingsen et al. (2010) findings can be reconciled with models of psychological game theory if we also recognize, in addition, the surprise-seeking motive. For their overall sample, they find

[10] For the relevant references, see Dhami et al., (2017).

[11] This design is not subject to other confounding influences. For instance, pre-play communication may enhance first and second order beliefs (Charness and Dufwenberg, 2006). Yet pre-play communication might influence actions not because players suffer from guilt-aversion, but rather because they may have a preference for promise-keeping (Vanberg, 2008).

that the correlation between second order beliefs and actions is not significantly different from zero (as in Ellingsen et al., 2010), but the situation is different at the individual level. When psychological factors are statistically significant, about 70% of the dictators are guilt-averse and about 30% are surprise-seeking. However, the behavior of the two types of players cancels out in the aggregate, giving rise to the appearance that there is no guilt-aversion.

Dhami et al. (2018) consider a public goods game, which has an explicit strategic interaction component. They extend the theoretical framework of psychological games to a two-player public goods game that takes account of guilt-aversion, surprise-seeking, attribution of intentions, and reciprocity. In an induced beliefs design, they find that all these emotions that underpin human morality, play an important role in explaining contributions in public goods games. In particular, guilt–aversion is, by far, the predominant finding at the level of the individuals, and for the aggregate data. They find that, for at least 30% of the subjects, the *attribution of intentions* behind guilt-aversion/surprise-seeking is statistically significant, although they cannot rule out this motive for the remaining subjects.

10. Markets, incentives, and morality

This section studies the relation between markets and human sociality. Consider first a few examples. Framing interactions between subjects in terms of market terminology (e.g., sellers, buyers, bargaining) can diminish moral considerations (Cappelen et al., 2013). When workers in a field experiment are given an in-kind gift (a water thermos) relative to an equivalent monetary gift of $20, their effort level increases by 25%, even when they have no preference for the thermos over $20 (Kube et al., 2012). Implicit incentives, such as bonuses, may highlight the moral aspects of ones actions, while explicit incentives, such as performance based pay, may turn-off the moral frame (Bowles and Polanía-Reys, 2010), or trigger *moral disengagement* (Bandura, 1991). When dictators in dictator game experiments are made to earn their endowments, or similar entitlements are created, they offer lower amounts to the receiver (Schotter et al., 1996; Hoffman et al., 2008). Experiments on rural communities in Columbia, who live on the edge of a forest, show that they are more likely to conserve common-resources if presented with the problem in terms of *local cooperative effort*, rather than *quota-based government regulation* (Cardenas et al., 2000).

Economists stress the role of *extrinsic motivation* that responds to external incentives, sometimes simply known as economic incentives. Yet, increasingly, behavioral economics has highlighted the role of *intrinsic motivation* (clearly, in addition to extrinsic motivation), and provided persuasive theoretical frameworks to study its effect on economic behavior (Bénabou and Tirole, 2003, 2006). Individuals signal to themselves, and to others, through costly actions, such as charitable giving, that they are good, moral, people. Indeed, extrinsic incentives, by removing or reducing the opportunity to engage in such signaling, may even crowd out intrinsic motivation. Ariely et al. (2009) found that charitable donations are lower when they are publicly announced and incentives are given for donations, presumably because they prevent the possibility of signalling and maintaining a positive self-image.

In an early and pioneering study, Titmus (1971) found that individuals are more likely to donate blood when they do so voluntarily through intrinsic motivation, rather than in the presence of monetary incentives. Using Swedish data, Mellström and Johannesson (2008) found strong gender effects of incentives for blood donation; a crowding-out effect for women but not for men. In contrast, using data from the American Red Cross blood drives (single events that solicit blood donation), Lacetera et al. (2012) found that incentives crowd-in blood donations. However, a significant increase in donations from incentives, in this study, came from substitution effects arising from other spatially and temporally separated blood drives. Goette and Stutzer (2008) found no effects of incentives on blood donations for long-term committed donors.

Gneezy and Rustichini (2000) found that when fines were levied for late arriving parents in private day care centres in Haifa, they arrived to pick up their children even later. This occurs because the fine places an extrinsic value on late arrivals, substituting for the intrinsic motivation of parents to not delay the carers. Parents continued to arrive late, even when the fine was removed, suggesting long-lasting, and negative, effects on intrinsic motivation. A similar interpretation may be given to the finding of Holmås et al. (2010): fines for overstaying in hospitals in Norway induced people to overstay even longer. Swiss residents were found to be more likely to agree to nuclear waste disposal in their communities when an appeal was made to their civic values (intrinsic motivation), as compared to being offered monetary compensation (Frey and Oberholzer-Gee, 1997).

Falk and Szech (2013) conduct an interesting experiment in which individuals had a choice between taking 10 euros or saving the life of a young healthy mouse who might be expected to live for 2 more years. In the non-market condition, subjects made this choice as isolated decision makers. In the market condition, subjects could bargain with another subject (bilateral setting) or several other subjects (multilateral setting). If subjects successfully bargain, they get 10 euros, but the life of the mouse is lost. Otherwise (unsuccessful bargaining), they lose 10 euros, but the mouse is saved. The percentage of subjects who are willing to accept 10 euros rather than save the life of the mouse in, respectively, the non-market, bilateral, and the multilateral treatments is, 45.9, 72.2 and 75.9. Thus, one is led to the conclusion that markets reduce morality.

The authors conjecture that markets may reduced morality in their experiments for the following 3 reasons, although the experiments cannot disentangle the relative contribution of each of the reasons. (1) Relative to the non-market treatment, the responsibility for killing the mouse is spread over greater number of people in the market treatments. This may reduce the guilt from killing the mouse in the market treatments. (2) In the process of bargaining with others, one may observe that others are willing to trade, hence, condemning the mous8; this might loosen one's own morality. (3) Markets might draw attention to a non-moral frame by focussing on bargaining, negotiations, and competition.

A range of interesting economic issues involve negative externalities in which actions by one party cause harm or disutility to another party. The typical policy response, e.g., corrective taxes, ensures that parties internalize the private and social costs of their actions. However, if economic agents care directly for social responsibility, might they internalize negative externalities anyway? For instance, many corporations stress the idea of corporate social responsibility, which requires corporations to take account of the larger

social interest, even at a cost in terms of private profits. Consumers too are often willing to pay extra for socially responsible products that do not involve child labour or cruelty to animals, or that are made with greener, more expensive, technologies.

Bartling et al. (2015) design experiments to consider these issues in a market/non-market setting and in two different datasets drawn from Swiss and Chinese subjects. In the baseline condition, there are 6 firms, 5 consumers, and 5 third parties. Firms have a choice of producing either a costless product that causes negative externalities worth 60 units to the third parties; or a costly product that costs 10 units and causes no externalities to the third parties. Each firm can sell one unit to one consumer, and each firm chooses, independently, a price and a type of the product. Each product gives each customer an identical value of 50 units, and consumers enter the market sequentially.

If consumers and firms have no social responsibility, then consumers wish to buy the cheapest product and firms wish to maximize monetary profits. In this case, we should only observe the externality-causing product in the market, giving, for each product sold, a negative social surplus ($50-60=10$). However, in the presence of social responsibility, we may observe the exchange of the more costly externality-free product. The game is played over 24 rounds. By not revealing the ID of players over successive rounds and randomly rematching them in each round, reputational effects are minimized or eliminated.

The results are as follows. The baseline condition quickly stabilizes at 45% of the products being externality-free. As expected, these are sold at a higher price relative to the externality-causing product, but the price difference is lower than the extra production cost of these products. Thus, in equilibrium, both sellers and buyers share in the costs of being socially responsible in competitive markets. When more competition between the sellers is introduced, the price drops further, but social responsibility does not. However, when the cost of production of the externality-free product is raised (from 20% of the surplus to 80% of the surplus), the degree of social responsibility falls. The authors estimate that the utilities of players in their model are best described as a combination of material utility and utility for socially responsible products. When the same experiment is repeated in China, the share of the externality-free product stabilizes at a lower level of 16%, suggesting lower norms of socially responsible behavior.

When Swiss and Chinese subjects are both asked to play a non-market allocation game with similar payoffs as those that arise in the market game, then the outcomes are very similar. Thus, prosociality in both societies is similar. Furthermore, the frequency of choices in the non-market setting that mitigate the negative effects on third parties (the analogue of negative externalities in the market setting) is relatively higher. Hence, markets do appear to reduce ethicality, which is consistent with the findings of Falk and Szech (2013).

Bartling and Özdemir (2017) consider the possibility that firms may engage in an unethical business opportunity, on the grounds that if they did not, someone else will (the 'replacement excuse'). Whether the replacement excuse is exercised by subjects in the experiment depends on the norm for such excuses. If there are no such norms, then the replacement excuse is more likely to be exercised. However, if there exists a norms that such an excuse is immoral, then this excuse is not used. The importance of this work is to show that existing norms influence whether markets reduce or enhance ethicality, hence,

without studying the interaction effects between the two, we might get misleading results.

Evidence collected from cross-cultural variation in the outcomes of the ultimatum game suggests that the two main factors that enhanced human sociality were the following. (1) Market integration in the community, i.e., the predominance of buying/selling and working for a wage. (2) Degree of cooperation in production (e.g., whether production is carried out on an individual basis or in a team). Indeed, these two factors alone explained 66% of the variation in outcomes in the ultimatum game.

Thus, *markets enhance sociality, even if they might diminish ethicality.* By not making this important distinction, researchers risk drawing erroneous conclusions. Ethicality does appear to be influenced by norms for ethicality, but this begs the interesting question of why there are norms for some types of ethical behavior, but not others.

11. Cross-Country differences in honesty

Several studies with relatively low stakes do not find any statistically significant cross-country differences in honesty. This includes a coin flip study from 16 countries in which the outcome heads was rewarded with a chocolate (Pascual-Ezama et al., 2015) and a die rolling task over 20 trials in which the incentive to report honestly was 10 cents (Mann et al., 2016). In a sender-receiver game, Holm and Kawagoe (2010) do not find any average differences in honesty levels when they compare Swedish and Japanese subjects.

Hugh-Jones (2015) performs a coin-flip experiment in 15 countries with 1535 subjects, using members of managed online panels that are typically used by firms for market research. Subjects received a monetary incentive of either $3 or $5 for reporting heads in a private toss of a coin. They were also asked to guess the level of honesty of other subjects from their and other countries. Significant variation was found in the level of honesty across the countries. Richer countries were more honest, on average, as were countries with a greater percentage of Protestant subjects. However, the main correlate of honesty is pre 1950 GDP differences, but not differential growth in GDP since 1950. The beliefs of subjects about the honesty of others, in their country and other countries, were fairly inaccurate.

Gächter and Schulz (2016) aim to explain cross-country differences in honesty in terms of the underlying causes. They construct a PRV (prevalence of rule violations) index for 159 countries. The PRV index captures 3 kinds of rule violations: *political fraud* (using an index of political rights), *tax evasion* (proxied by the size of the shadow economy), and *corruption* (derived from the World Bank's control of corruption index). The authors use a sample of comparable student subjects from 23 countries, such that the distribution of PRV in the sample was representative of the original list of 159 countries. Subjects rolled a six-sided die, twice, in the die-in-a-cup method and were asked to report only the privately observed outcome of the first throw. Payments equalled the self-reported claim for numbers 1-5, and zero for number 6.

Figure 11.1 shows the cumulative density functions (CDFs) for the PRV data from the sample countries, separated into high PRV countries (darker CDFs) and low PRV countries (lighter CDFs). It also shows CDFs for various honesty benchmarks that we define below, and an inset histogram for the reported claims from high and low PRV countries.

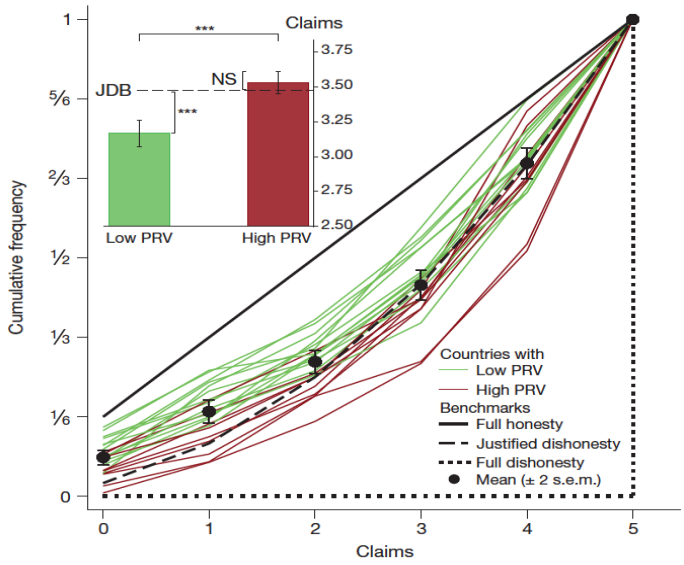


Figure 11.1: Cumulative distribution functions (CDFs) of self-reported outcomes relative to various honesty benchmarks. Source: Gächter and Schulz (2016)

The expected claim under full-honesty, i.e., when the claim equals the observed outcome, is 2.5; in the inset histogram, this is the benchmark against which reported claims are measured. The CDF under the full-honesty benchmark, where each outcome occurs with a probability 1/6, is shown as a diagonal line. The CDF under full-dishonesty (always report an outcome of 5) is shown as the dotted line. The authors also construct the CDF under *justified ethicality*, which relies on the idea that individuals feel it is less immoral to report the "higher of two numbers on the two throws of the die", rather than lie outrightly; this is based on Shalvi et al. (2011). Under justified ethicality, a 6 (and a claim of 0) occurs in 1/36 cases; a 1 (and a claim of 1) occurs in 3/36 cases, i.e., for any of the outcomes (6,1), (1,6) or (1,1); a 2 (and a claim of 2) occurs in 5/36 cases; i.e., following the outcomes (6,2), (2,6), (1,2), (2,1), (2,2); and so on. The CDF corresponding to justified ethicality is shown as the dotted curve that passes through the middle of the other CDFs in Figure 11.1.

The results are as follows:

1. On average, the CDFs from all countries are neither close to the full-honesty benchmark, nor to the full-dishonesty benchmark. For each subject pool, using a Kolgomorov-Smirnov test, one can statistically reject the null of equality of the CDFs with the full-honesty benchmark. However, for 13 out of 23 subject pools, we cannot reject the equality of the observed CDFs with the CDF arising from the justified ethicality benchmark.

2. Under full-honesty, the expected claim should be 2.5. Subjects in high PRV countries claim significantly more relative to the low PRV countries (3.53 versus 3.17; $t = 5.84$, $p < 0.001$).

3. The fraction of income maximizers (those who report a 5) is not significantly correlated with the PRV index. However, the fraction of individuals who are always honest (percentage of those who report a 6, multiplied by 6) varies from 4.3% to 87%.

4. No gender effects are found.

12. Honesty and neuroeconomics"

Garrett et al. (2016) examine changes in dishonesty when the opportunity to be dishonest is repeated over time. In their setup, an *advisor* is asked to provide advice to an uninformed *estimator* about how much money there is in a jar. The advisor's payoff is increasing in the amount of money that he/she states in this advice. When this game is repeated, and only the advisor benefits from the dishonesty (self-serving dishonesty), the advisor is found to engage in ever increasing levels of dishonesty. The authors term this phenomenon as the *slippery slope of dishonesty*. However, when dishonesty is other-serving, i.e., it benefits the estimator but not the advisor, this phenomenon is not observed. Hence, it is selfishness, not altruism, that gives rise to the slippery slope.

The really innovative feature of the study is that the authors discover a continually reduced BOLD signal in the amygdala as an advisor engages in the same level of dishonesty, conditional on having been dishonest in the past. The authors interpret this as reduced guilt from successive acts of being dishonest. In the context of this experiment, Engelmann and Fehr (2016) suggest examining the interaction of the amygdala signal with other interconnected brain areas that are involved with dishonest actions.

In a die-throwing task, Maréchal et al. (2017) apply *transcranial direct current stimulation* (tDCS) to the right dorsolateral prefrontal cortex (rDLPFC), a region of the brain, suspected to resolve conflicts between personal gains and honesty. The stakes were relatively high; subjects could earn 90 Swiss Francs in the experiment if they tried to maximize their monetary earnings. In the sham condition, which acts as a baseline, 37% are dishonest. Applying anodal tDCS to the rDLPFC (to enhance neural excitability), as expected, the level of honesty increased. However, when they apply cathodal tDCS to rDLPFC (to decrease neural excitability), then there was no appreciable effect on honesty. Interestingly, tDCS did not affect the amount of money that was kept by dictators in a dictator game. Hence, the increased honesty observed in the die task, on account of anodal tDCS, is unlikely to have been caused by reduced material interest.

Green and Paxton (2009) found that when deliberating between a honest and a dishonest option, honest people do not exhibit significantly more brain activity in areas of the brain that are associated with conflict and cognitive control. This suggests that these individuals did not face a temptation to cheat, rather they prefer to be intrinsically honest.

13. Conclusions

The homo-economicus model is not supported by the empirical evidence. A rich body of evidence now provides strong support for a homo-behavioralis model. Exclusive reliance

on the homo-economicus model is neither justified on empirical grounds, nor on the mistaken belief that it leads to a more parsimonious account of real phenomena. Indeed, in order to ensure consistency with the data, several auxiliary conditions must be invoked in neoclassical models to account for the missing motivations found in homo-behavioralis–a sort of 'missing variables bias' in the theoretical models. This does not necessarily make the homo-economicus model simpler or more parsimonious. The non-compliance with the empirical evidence is sufficient grounds for moving beyond the homo-economicus model. The main aim of this paper is to make a strong case for such a move. While we now have much better data on the richness of human morality, the impact on theoretical models within behavioral economics, despite commendable progress, is likely to be felt in the future.

14. References

1. Abeler, J., Becker, A. and Falk, A. (2014) Representative evidence on lying costs. Journal of Public Economics 113: 96–104.

2. Abeler, J. Nosenzo, D. and Raymond, C. (2016) Preferences for truth-telling. CEDEX Discussion Paper No. 2016-13.

3. Akerlof, G.A., and Kranton, R.E. (2005). Identity and the economics of organizations. Journal of Economic Perspectives. 19(1): 9–32.

4. al-Nowaihi, A. and Dhami, S. (2015). Evidential equilibria: Heuristics and biases in static games of complete information. Games. 6: 637-676.

5. Allport, G. W. (1955). Becoming: Basic considerations for a psychology of personality. New Haven, CT: Yale University Press

6. Alm, J., Jackson, B. R. and Mckee, M. (2009) Getting the Word Out: Enforcement Information Dissemination and Compliance Behavior. Journal of Public Economics. 93(3-4): 392-402.

7. Alm, J., Bloomquist, K.M., McKee, M. (2015). On the external validity of laboratory tax compliance experiments. Economic Inquiry 53(2): 1170–1186.

8. Andreoni, James (1990). Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. Economic Journal. 100 (401): 464–477.

9. Andreoni, J., and Miller, J.H. (2002). Giving according to GARP: an experimental test of the consistency of preferences for altruism. Econometrica. 70(2): 737–753.

10. Ariely, D., Bracha, A., and Meier, S. (2009). Doing good or going well? Image motivation and monetary incentives in behaving prosocially. American Economic Review. 99(1): 544–555.

11. Armantier, O., Boly, A. (2013). Comparing corruption in the laboratory and in the field in Burkina Faso and in Canada. The Economic Journal 123: 1168-1187.

12. Azar, O.H., Yosef, S. and Bar-Eli, M. (2013) Do customers return excessive change in a restaurant? A field experiment on dishonesty. Journal of Economic Behavior and Organization. 93: 219–226.

13. Balafoutas, L., Czermak, S., Eulerich, M., and Fornwagner, H. (2017). Incentives for dishonesty: An experimental study with internal auditors. Working Papers in Economics and Statistics, University of Innsbruck. No. 2017-06.

14. Bandura, A. (1991). Social congnitive theory of moral thought and action. In: W.M. Kurtines, J. Gewirtz and J.L. Lamb (eds), Handbook of Moral Behavior and Development: Volume I, Theory. Hillsdale, New Jersey: Lawrence Erlbaum and Associates, pp. 45–103.

15. Barkan, R., Ayal, S., Gino, F., & Ariely, D. (2012). The pot calling the kettle black: Distancing response to ethical dissonance. Journal of Experimental Psychology: General, 141, 757–773.

16. Bartling, B., Engl, F, Weber, R. A. (2014) Does Willful Ignorance Deflect Punishment? – An Experimental Study. European Economic Review. 70: 512-524.

17. Bartling, B., and Fischbacher, U. (2012). Shifting the blame: on delegation and responsibility. Review of Economic Studies. 79(1): 67–87.

18. Bartling, B., and Özdemir, Y. (2017) The Limits to Moral Erosion in Markets: Social Norms and the Replacement Excuse. CESifo Working papers, Vol 17, No. 93.

19. Bartling, B., Weber, R., and Yao, L. (2015). Do Markets Erode Social Responsibility? Quarterly Journal of Economics, 130(1): 219–66.

20. Bateson, M., Nettle, D., and Roberts, G. (2006). Cues of Being Watched Enhance Cooperation in a Real-World Setting. Biology Letters. 2: 412–14.

21. Battigalli, P., and Dufwenberg, M. (2007). Guilt in games. American Economic Review. 97(2): 170-176.

22. Battigalli, P., and Dufwenberg, M. (2009). Dynamic psychological games. Journal of Economic Theory. 144(1): 1-35.

23. Bénabou R (2013) Groupthink: Collective delusions in organizations and markets. Review of Economic Studies. 80(2):429–462.

24. Bénabou, R., and Tirole, J. (2003). Intrinsic and extrinsic motivation. Review of Economic Studies. 70(3): 489–520.

25. Bénabou R, Tirole J (2006) Incentives and prosocial behavior. American Economic Review. 96(5):1652–1678.

26. Bicchieri, C. (2006) The Grammar of Society: The Nature and Dynamics of Social Norms. Cambridge University Press: Cambridge.

27. Blount, S. (1995) When Social Outcomes Aren't Fair: The Effect of Causal Attribution on Preferences. Organizational Behavior and Human Decision Processes, 63(2): 131-44

28. Bowles, S., and Polania-Reyes, S. (2012). Economic incentives and social preferences: substitutes or complements? Journal of Economic Literature. 50(2): 368–425.

29. Breton, A. and Ronald Wintrobe, R. (1978) A Theory of 'Moral' Suasion. The Canadian Journal of Economics. 11(2): 210-219.

30. Camerer, C.F. (2003). Behavioral game theory: Experiments in strategic interaction. Princeton University Press: Princeton.

31. Camerer, C.F. (2015). The Promise and Success of Lab-Field Generalizability in Experimental Economics: A Critical Reply to Levitt and List. in Fréchette, G.R., and Schotter, A. (eds.) Handbook of Experimental Economic Methodology. Oxford University Press: Oxford. pp. 249-295.

32. Cappelen, A.W., Sørensen E.Ø., and Tungodden, B. (2013). When do we lie? Journal of Economic Behavior and Organization. 93: 258–265.

33. Cardenas, J.C., Stranlund, J.K., and Willis, C.E. (2000). Local environmental control and institutional crowding-out. World Development. 28(10): 1719–1733.

34. Carpenter, J., Matthews, P. H., Schirm, J., 2010. Tournaments and office politics: Evidence from a real effort experiment. The American Economic Review, 100(1), 504-517.

35. Charness, G., and Dufwenberg, M. (2006). Promises and partnership. Econometrica. 74(6): 1579–1601.

36. Charness, G., Masclet, D., Villeval, M. C., 2014. The Dark Side of Competition for Status. Management Science, 60(1), 38-55.

37. Cohn, A., Fehr, E. and Mare´chal, M.A. (2014) Business culture and dishonesty in the banking industry. Nature 516: 86–89.

38. Cohn, A., Maréchal, M. and Noll, T. (2013) Bad Boys: How Criminal Identity Salience Affects Rule Violation. The Review of Economic Studies. 82(4): 1289–1308.

39. Conrads, J., Irlenbusch, B., Rilke, R. M., Walkowitz, G., 2013. Lying and team incentives. Journal of Economic Psychology, 34, 1-7.

40. Dai, Z., Galeotti, F., Villeval, M. C. (2017) Cheating in the Lab Predicts Fraud in the Field: An Experiment in Public Transportation, forthcoming Management Science.

41. Dana, J., Cain, D. M., & Dawes, R. M. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. Organizational Behavior and Human Decision Processes, 100: 193–201.

42. Dana, J., Weber, R.A., and Kuang, J.X. (2007). Exploiting moral wriggle room: experiments demonstrating an illusory preference for fairness. Economic Theory. 33: 67–80.

43. Dhami, S., and al-Nowaihi, A. (2010a). Existence of a Condorcet winner when voters have other-regarding preferences. Journal of Public Economic Theory. 12(5): 897–922.

44. Dhami, S., and al-Nowaihi, A. (2010b). Redistributive policy with heterogeneous social preferences of voters. European Economic Review. 54(6): 743–759.

45. Dhami, S. (2016) The foundations of behavioral economic analysis. Oxford University Press: Oxford.

46. Dhami, S., Wei, M., and al-Nowaihi, A. (2018) Public goods games and psychological utility: Theory and evidence. Forthcoming in Journal of Economic Behavior and Organization.

47. Dreber, A. and Johannesson, M. (2008) Gender differences in deception. Economics Letters 99(1): 197–199.

48. Dufwenberg, M., Heidhues, P., Kirchsteiger, G., Riedel, F., et al. (2011). Other-regarding preferences in general equilibrium. Review of Economic Studies. 78(2): 613–639.

49. Dufwenberg, M., and Kirchsteiger, G. (2004). A theory of sequential reciprocity. Games and Economic Behavior. 47(2): 268-298.

50. Eckel, C.C., and Petrie, R. (2011). Face value. American Economic Review. 101(4): 1497–1513.

51. Ellingsen, T., and Johannesson, M. (2004). Promises, threats and fairness. Economic Journal. 114(495): 397–420.

52. Ellingsen, T., M. Johannesson, S. Tjøtta, and G. Torsvik (2010) Testing Guilt Aversion. Games and Economic Behavior. 68: 95-107.

53. Elster, J. (1998) Emotions in Economic Theory. Journal of Economic Literature. 36: 47-74.

54. Engelmann, J., B. and Fehr, E. (2016) The slippery slope of dishonesty. Nature Neuroscience. 19: 1543-1544.

55. Erat, S. (2013) Avoiding lying: The case of delegated deception. Journal of Economic Behavior & Organization. 93: 273-78

56. Erat, S. and Gneezy, U. (2012) White lies. Management Science 58(4): 723–733.

57. Falk, A., and Szech, N. (2013). Morals and markets. Science. 340(6133): 707–711.

58. Faravelli, M., Friesen, L., Gangadharan, L., 2015. Selection, tournaments, and dishonesty. Journal of Economic Behavior Organization, 110, 160-175.

59. Fehr, E., and Gächter, S. (2000). Cooperation and punishment in public goods experiments. American Economic Review. 90(4): 980-994.

60. Fehr, E., and Schmidt, K. (2006) The economics of fairness, reciprocity and altruism: Experimental evidence and new theories. in Serge-Christophe Kolm and Jean Mercier Ythier (eds.) Handbook of the Economics of Giving, Altruism and Reciprocity, Volume 1., Elsevier.

61. Fehr, E., and Schneider, F. (2010). Eyes are on us, but nobody cares: are eye cues relevant for strong reciprocity? Proceedings of the Royal Society B: Biological Sciences, 277: 1315–1323.

62. Fischbacher, U., and Föllmi-Heusi, F. (2013). Lies in disguise: an experimental study on cheating. Journal of the European Economic Association. 11(3): 525–547.

63. Fosgaard, Toke Reinholt, Lars Gaarn Hansen, and Marco Piovesan. (2013) Separating Will from Grace: An Experiment on Conformity and Awareness in Cheating.Journal of Economic Behavior & Organization 93: 279-84.

64. Fréchette, G.R., and Schotter, A. (eds.) Handbook of Experimental Economic Methodology. Oxford University Press: Oxford.

65. Frey, B.S., and Oberholzer-Gee, F. (1997). The cost of price incentives: an empirical analysis of motivation crowding-out. American Economic Review. 87(4): 746–755.

66. Friesen, Lana, and Lata Gangadharan. (2012) Individual Level Evidence of Dishonesty and the Gender Effect. Economics Letters 117, 3: 624-26.

67. Gächter, S. and Schulz, J.F. (2016) Intrinsic honesty and the prevalence of rule violations across societies. Nature 531: 496–499.

68. Garbarinoa, E., Slonim. R., Marie Claire Villeval, M. C. (2017) Loss Aversion and lying behavior: Theory, estimation and empirical evidence. mimeo.

69. Garrett, N., Lazzaro, S.C., Ariely, D. & Sharot, T. (2016) The brain adapts to dishonesty. Nature Neuroscience. 19: 1727–1732.

70. Geanakoplos, J., Pearce, D., and Stacchetti, E. (1989). Psychological games and sequential rationality. Games and Economic Behavior. 1(1): 60-79.

71. Gibson, R., Tanner, C., Wagner, A.F., 2013. Preferences for truthfulness: heterogeneity among and within individuals. American Economic Review 103 (1): 532–548.

72. Gintis, H. (2017). Individuality and Entanglement: The Moral and Material Bases of Social Life. Princeton, NJ: Princeton University Press.

73. Gintis, H. (2009). The bounds of reason: Game theory and the unification of the social sciences. Princeton, NJ: Princeton University Press.

74. Gintis, H. and Khurana, R. (2016) Corporate Corruption and the Failure of Business School Education. mimeo Harvard University.

75. Gino, F. and Ariely, D. (2012) The dark side of creativity: Original thinkers can be more dishonest. Journal of Personality and Social Psychology 102(3): 445–459.

76. Gino F, Ayal S, Ariely D. (2013) Self-Serving Altruism? The Lure of Unethical Actions that Benefit Others. Journal of Economic Behavior and Organization. 93: 285-292.

77. Gino, F., Krupka, E., & Weber, R. (2013). License to cheat: Voluntary regulation and ethical behavior. Management Science, 59(10), 2187-2203.

78. Gino F, and Pierce L. (2010) Lying to level the playing field: Why people may dishonestly help or hurt others to create equity. Journal of Business Ethics. 95(1):89–10.

79. Gino, F., Schweitzer, M.E., Mead, N., Ariely, D., (2011). Unable to resist temptation: how self-control depletion promotes unethical behavior. Organizational Behavior and Human Decision Processes 115, 191–203.

80. Gneezy, U. (2005) Deception: The role of consequences. American Economic Review 95(1), 384–394.

81. Gneezy, U., Gravert, C., Saccardo, S., Tausch, F. (2017) A Must Lie Situation: Avoiding Giving Negative Feedback. Games and Economic Behavior. 102: 445-454.

82. Gneezy, U., Rockenbach, B. and Serra-Garcia, M. (2013) Measuring lying aversion. Journal of Economic Behavior and Organization 93: 293–300.

83. Gneezy, U., and Rustichini, A. (2000). A fine is a price. Journal of Legal Studies. 29(1): 1–17.

84. Goette, L.F., and Stutzer, A. (2008) Blood donations and incentives: evidence from a field experiment. IZA, Discussion Paper 3580.

85. Greene JD, Paxton JM (2009) Patterns of neural activity associated with honest and dishonest moral decisions. PNAS 106:12506–12511.

86. Grolleau, G., Kocher, M. G. and Sutan, A. (2016). Cheating and loss aversion: do people lie more to avoid a loss? Forthcoming in Management Science. 62(12): 3428–3438.

87. Haley, K.J., and Fessler, D.M.T. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous economic game. Evolution and Human Behavior. 26(3): 245–256.

88. Gneezy, U. Alex Imas, Kristóf Madarász (2014) Conscience Accounting: Emotion Dynamics and Social Behavior. Management Science 60(11):2645-2658.

89. Hallsworth, M., List, J.A., Metcalfe, R. D., Vlaev, I. (2017) The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. Journal of Public Economics 148, 14- 31, 2017

90. Hanna, R. and Shing-Yi Wang, S.-Y. (2017) Dishonesty and Selection into Public Service: Evidence from India. American Economic Journal: Economic Policy, 9(3): 262–290

91. Harbaugh, W; Mayr, U; Burghart, D (2007). Neural Responses to Taxation and Voluntary Giving Reveal Motives for Charitable Donations. Science. 316 (5831): 1622–1625.

92. Hays, Chelsea, and Leslie J. Carver. (2014) Follow the Liar: The Effects of Adult Lies on Children's Honesty. Developmental Science 17, 6: 977-83.

93. Herrmann, B., Thöni, C., and Gächter, S. (2008). Antisocial punishment across societies. Science. 319(5868): 1362–1367.

94. Hoffman, E., McCabe, K., and Smith, V.L. (2008). Preferences and property rights in ultimatum and dictator games. In: C.R. Plott and V.L. Smith (eds), Handbook of Experimental Economics Results, Volume 1. Amsterdam: North-Holland, pp. 417–422.

95. Holm, H.J. and Kawagoe, T. (2010) Face-to-face lying – An experimental study in Sweden and Japan. Journal of Economic Psychology 31(3): 310–321.

96. Holmås, T.H., Kjerstad, E., Lurås, H., and Straume, O.R. (2010). Does monetary punishment crowd out pro-social motivation? A natural experiment on hospital length of stay. Journal of Economic Behavior and Organization. 75(2): 261–267.

97. Houser, D., Vetter, S. and Winter, J. (2012) Fairness and cheating. European Economic Review 56(8): 1645– 1655.

98. Houser, D., List, J.A., Piovesan, M., Samek, A. and Winter, J. (2016) Dishonesty: From parents to children. European Economic Review 82: 242–254.

99. Hugh-Jones, D. (2015) Honesty and beliefs about honesty in 15 countries. Mimeo University of East Anglia.

100. Hurkens, S. and Kartik, N. (2009). Would I Lie to You? On Social Preferences and Lying Aversion. Experimental Economics, 12(2): 180-92.

101. Kajackaite, A. and Gneezy, U. (2017) Incentives and Cheating. Games and Economic Behavior. 102: 433–444.

102. Kahneman, D. and Tversky, A. (1979) Prospect Theory: An Analysis of Decision under Risk. Econometrica, 47(2): 263-291.

103. Kartik, N. (2009). Strategic communication with lying costs. Review of Economic Studies. 76(4): 1359–1395.

104. Khalmetski, K., Ockenfels, A., and Werner, P. (2015). Surprising Gifts. Journal of Economic Theory. 159: 163-208.

105. Kocher, M. G. and Schudy, S. (2017) I Lie? We Lie! Why? Experimental Evidence on a Dishonesty Shift in Groups. Management Science. Published online in Articles in Advance 03 Aug 2017.

106. Kocher M, Strauß S, Sutter M (2006) Individual or team decision making-Causes and consequences of self-selection. Games and Economic Behavior 56(2): 259–270.

107. Kube, S., Maréchal, M.A., and Puppe, C. (2012). The currency of reciprocity: gift exchange in the workplace. American Economic Review. 102(4): 1644–1662.

108. Lacetera, N., Macis, M., and Slonim, R. (2012). Will there be blood? Incentives and displacement effects in pro-social behaviour. American Economic Journal: Economic Policy. 4(1): 186–223.

109. Leibbrandt, A. Pushkar Maitra and Ananta Neelim (2017) Large Stakes and Little Honesty? Experimental Evidence from a Developing Country. Monash Business School, Department of Economics Discussion Paper No. 13/17.

110. Levitt, S.D., and List, J.A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? Journal of Economic Perspectives. 21(2):153–174.

111. Mann, H. E., Garcia-Rada, X., Hornuf, L., Tafurt, J., and Ariely, D. (2016) Cut from the Same Cloth: Similarly Dishonest Individuals Across Countries.Journal of Cross-Cultural Psychology 47(6): 858–874.

112. Maréchal, A., Cohn, A., Ugazio, G. and Ruff, C. C (2017) Increasing honesty in humans with noninvasive brain stimulation. PNAS. 114(17): 4360–4364.

113. Mazar N, Ariely D. (2006) Dishonesty in everyday life and its policy implications. Journal of Public Policy and Marketing. 25(1):117–126.

114. Mazar N, Amir O, Ariely D. (2008) The dishonesty of honest people: A theory of self-concept maintenance. Journal of Marketing Research. 45:633–644.

115. Mead, N., Roy, F., Baumeister, F.G., Schweitzer, F E.M., Ariely, D. (2009). Too tired to tell the truth: self-control resource depletion and dishonesty. Journal of Experimental Social Psychology. 45(3), 594–597.

116. Mullainathan, S. and Bertrand, M. (2004) Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. The American Economic Review. 94(4): 991-1013.

117. Pascual-Ezama, D., Fosgaard, T., Cardenas, R. et al. (2015) Context dependent cheating: Experimental evidence from 16 countries. Journal of Economic Behavior & Organization. 116: 379–386.

118. Pe'er, E., Acquisti, A., & Shalvi, S. (2014). "I cheated, but only a little": Partial confessions to unethical behavior. Journal of Personality and Social Psychology. 106: 202–217

119. Ploner, M. and Regner, T. (2013) Self-image and moral balancing: An experimental analysis. Journal of Economic Behavior & Organization. 93: 374– 383.

120. Pruckner, G.J., and Sausgruber, R. (2013). Honesty on the streets: a field study on newspaper purchasing. Journal of the European Economic Association. 11(3): 661–679.

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

122. Reuben, E. and Stephenson, M. (2013) Nobody likes a rat: On the willingness to report lies and the consequences thereof. Journal of Economic Behavior & Organization, 93 (issue C): 384-391.

123. Rosenberg, M. (1979). Conceiving the self. New York, NY: Basic Books.

124. Ross, L., Greene, D., and House, P. (1977). The 'false consensus effect': an egocentric bias in social perception and attribution processes. Journal of Experimental Social Psychology. 13(3): 279-301.

125. Schindler, S. and Pfattheicher, S. (2017) The frame of the game: Loss-framing increases dishonest behavior. Journal of Experimental Social Psychology. 69: 172–177.

126. Schotter, A., Weiss, A., and Zapater, I. (1996). Fairness and survival in ultimatum and dictatorship games. Journal of Economic Behavior and Organization. 31(1): 37–56.

127. Schweitzer M.E., Ordóñez, L., Douma, B. (2004) Goal setting as a motivator of unethical behavior. Academy of Management Journal. 47: 422–432.

128. Shalvi S (2012) Dishonestly increasing the likelihood of winning. Judgment Decision Making. 7:292–303.

129. Shalvi, S., Dana, J., Handgraaf, M.J.J., De Dreu, C.K.W. (2011) Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. Organizational Behavior and Human Decision Processes. 115:181–190.

130. Shalvi, S., Eldar, O., and Bereby-Meyer, Y. (2012). Honesty requires time (and lack of justifications). Psychological Science. 23: 1264–1270.

131. Shalvi, S., & Leiser, D. (2013). Moral firmness. Journal of Economic Behavior & Organization. 93: 400–407.

132. Sutter, M., 2009. Deception through telling the truth?! Experimental evidence from individuals and teams. Economic Journal. 119: 47–60.

133. Tajfel, H., Billig, M.G., Bundy, R.P., and Flament, C. (1971). Social categorization and inter-group behavior. European Journal of Social Psychology. 1(2): 149–178.

134. Tajfel, H., and Turner, J. (1986). The social identity theory of intergroup behavior. In: W.G. Austin and S. Worchel (eds), The Psychology of Intergroup Relations. Chicago: Nelson-Hall, pp. 7–24.

135. Titmuss, R.M. (1971). The Gift Relationship: From Human Blood to Social Policy. New York: Pantheon Books.

136. Utikal, V. and Fischbacher, U. (2013) Disadvantageous lies in individual decisions. Journal of Economic Behavior and Organization. 85(1): 108–111.

137. van derWeele, J. J., Kulisa, J., Kosfeld, M. and Friebel, G.. (2014). Resisting Moral Wiggle Room: How Robust Is Reciprocal Behavior? American Economic Journal: Microeconomics. 6(3): 256-64.

138. Vanberg, C. (2008). Why do people keep their promises? An experimental test of two explanations. Econometrica. 76(6): 1476-1480.

139. Wiltermuth, S. S. (2011). Cheating more when the spoils are split. Organizational Behavior and Human Decision Processes. 115: 157–168.