Is Dishonesty Contagious?

By Robert Innes[A] and Arnab Mitra[B]

Abstract

When an individual believes that peers are predominantly untruthful in a given situation, is he/she more likely to be untruthful in that situation? We study this question in a deception experiment patterned after Gneezy (2005), finding evidence that dishonesty is indeed contagious. We argue that existing theories of other-regarding preferences are unlikely to explain this result.

1. Introduction

The importance of individual honesty and trustworthiness in economic interactions is well known. These attributes facilitate cooperative relationships, enable contracts, strengthen legal and regulatory institutions, and as a result, promote economic growth (Zak and Knack, 2001; Guiso, Sapienza and Zingales, 2004). Also well known are vast differences in these attributes across cultures and countries. Figure 1 illustrates these differences, showing proportions of world population and world economic activity (respectively) that derive from countries with high, medium, and low levels of corruption, as measured by Transparency International's 2005 corruption perception index (CPI). Without reading too much into these coarse numbers (which, of course, raise complex questions of cause and effect), we note a stylized fact: The distribution of corruption is largely bi-modal, with the vast majority of both population and economic activity in either the low CPI (advanced developed) or high CPI (Third World and "transition") countries.

In this paper, we explore a possible contributing explanation for this "fact" that is rooted in individual preferences.[1] Specifically, we conjecture that honesty is

---

[A] University of California, Merced.
[B] University of Arizona.
[1] Another possible explanation for this "fact" is that there is a vicious cycle in which low incomes promote corruption which, in turn, deters growth and so on. There is a vast literature on the evolution of institutions and their relationship to corruption and growth (see, for example, Acemoglu, et al,

contagious in the following sense: If a majority of one's peers are honest, an individual is more likely to suffer an aversion penalty / disutility when behaving dishonestly. If so, honesty breeds honesty and dishonesty breeds dishonesty.

We study this conjecture in the context of a simple deception experiment, wherein we stimulate different subject perceptions of the propensity for honesty in the overall group of experimental subjects. We then examine the resulting impact on an individual's choice between truthful and untruthful behavior. Our experiments mimic the original deception game designed by Uri Gneezy (2005), who studied the effects of different payoffs on individuals' aversion to untruthful behavior.[2] Unlike Gneezy (2005), we consider a single set of payoffs in each experiment and focus on the possibility of contagion. Because a central motive for our inquiry is to determine whether a perceived norm of honesty can spur more truthful conduct in a society that is considered corrupt, we conduct our experiments in both a low CPI country (Arizona, USA) and a high CPI country (Calcutta, India). In doing so, we find evidence for contagion in two senses: (1) the perception of a strong group propensity for dishonesty promotes untruthful behavior when subjects are otherwise predominantly honest (our Arizona experiment and survey); and (2) the perception of a strong group propensity for honesty promotes truthful behavior when subjects are otherwise predominantly untruthful (our India experiment).

To our knowledge, the only study that (indirectly) addresses the question of contagion in honesty is Fisman and Miguel's (2007) famous paper on the tendency for diplomats to garner parking tickets in New York; they find that the immunity-protected foreigners take their home country propensities for lawlessness with them. While these results might be interpreted as evidence against contagion (because diplomats seem to ignore U.S. values in their behavior), we believe that such

2001). One interpretation of our paper, in the context of this literature, is that, from the level of individual preferences, there may be some self-reinforcing dynamics to the evolution of bad and good economic institutions.

[2] Other recent experimental work on deception games include Sanchez-Pages and Vorsatz (2006), who study links between a subject's willingness to punish lies of others and their aversion to lying; Hurkens and Kartik (2009), who elaborate on Gneezy's (2005) results; Ederer and Fehr (2007), who study impacts of deception and aversion to lying in a principal agent game; Sutter (2009), who identifies sophisticated deception; Rode (2008), who studies effects of competitive and cooperative priming on subjects' honesty; and Charness and Dufwenberg (2005) who provide an alternative (guilt aversion) interpretation of Gneezy's (2005) findings.

inferences are misplaced for two reasons. First, there is no ceteris paribus in this comparison; diplomats may well temper their lawless behavior, relative to what they would do if protected by immunity in their home countries. Second, the empirical observation may be a reflection of different relevant peer groups for diplomats from different countries, consistent with the contagion hypothesis. We therefore offer a direct test of contagion in this paper.

Our study is related to recent experimental economics literature on the effects of social information on behavior and a large psychology literature on conformity (see Cialdini and Goldstein, 2004, for a review), but is distinguished from this work by its focus on subjects' truthfulness – our research question. Unlike a significant subset of the literature (but not all), our design also voids prevalent theoretical explanations for "conformity" and obeying social norms, including social sanctions for the violation of norms (see Fehr and Fischbacher, 2004), incentives to obtain social esteem (Akerlof, 1980; Bernheim, 1994), and benefits from others' information (Banerjee, 1992, Bikhchandaria, Hirshleifer and Welch, 1992). In our experiments, individual actions are unobservable to anyone other than the individual; there is no possibility of social sanction or building social esteem; and what others do has no bearing on the payoff consequences of individual decisions. In Section 3, we explore the scope for theories of social (other regarding) preferences to explain our findings, arguing that the behavior we observe is likely a symptom of hard-wired contagious preferences.[3]

Perhaps most closely related to the present paper is work on social information in dictator games. In Bicchieri and Xiao (2009), dictators are given different information about the proportion of subjects in a prior session who *were* "fair" vs. "selfish," and who believe dictators *should be* fair vs. selfish. Their results generally suggest that fairness in actions is contagious. Krupka and Weber (2009) expose dictators to a sample of four (fair vs. selfish) allocations of prior dictators and find a significant increase in the fraction of fair allocations when the sample share is

---

[3] In psychology, some conformist behavior is explained as "behavioral mimicry" (such as yawning) and/or "automatic activation" that provides "an adaptive shortcut that maximizes the likelihood of effective action with minimal expense of one's cognitive resources" (Cialdini and Goldstein, 2004, p. 609). We believe that our results are likely to be the outcome of such reflexive responses, as we discuss in Section 3.

(3/4) or 1, vs. (1/2) or less. Cason and Mui (1998) find that exposing dictators to one prior dictator allocation decision (vs. irrelevant information) reduces their propensity for selfish allocations. Duffy and Kornienko (2009) show that introducing a tournament that alternately ranks subjects' givings or earnings, significantly promotes generous and selfish allocations respectively; for our purposes, these results could be interpreted as dictators' response to a norm revealed by the choice of tournament. As in our experiments, subjects' actions in these studies are private; there are no social sanctions or rewards; and what others do is irrelevant to payoffs.[4]

The major difference in our experiments is the focus on deception rather than dictator games. This distinction, we believe, is central. Selfish behavior (as in a dictator game) and dishonest behavior (as in a deception game) are very different phenomena. While "fairness" may help promote cooperative relationships (like honesty), "selfish" and acquisitive impulses can promote effort and innovation that are at the core of a thriving market economy. In contrast, the negative consequences of dishonesty and corruption for economic prosperity are well documented. Perhaps for this reason, the culture and psychology of the two phenomena are also different. Selfishness is sometimes heralded as a symptom of the drive to compete and win, as in a game, but in other contexts, scorned as an impediment to cooperative relationships. Honesty, on the other hand, is consistently promoted as a value and virtue by church and community, suggesting that contagion may be less likely. Rode (2008), for example, finds that subjects' honesty is insensitive to competitive vs. cooperative priming. Given the importance of honesty to economic growth, and differences in norms of conduct around the world, it is also important to study contagion in honesty in different countries, as we do with our twin experiments in India and the U.S.. In sum, honesty (in deception games) and fairness (in dictator games) are different phenomena, and results from dictator experiments cannot be readily translated to deception experiments; indeed, even small framing differences in

---

[4] The possible exception is Duffy and Kornienko (2009), where winning a tournament, even if the winner is only identified anonymously (by ID number), may provide some intrinsic reward.

dictator games are known to have significant effects on behavior, as shown by List (2007) and Bardsley (2008).[5]

2. Experimental and Survey Evidence

A. The Arizona Classroom Experiment

To elicit honest or dishonest decisions from subjects, we closely follow the deception game designed by Uri Gneezy (2005). In this game, there are two possible payoff distributions for each pair of players, with each pair comprised of a "Sender" and a "Receiver." The two distributions are represented by Options A and B. Only the Sender is informed about the payments associated with the two options, one of which is advantageous to the Sender and the other of which is advantageous to the Receiver. The Sender sends one of two messages to the Receiver:

> Message A: "Option A will earn you (the Receiver) more money than Option B."

> Message B: "Option B will earn you (the Receiver) more money than Option A."

A message is *truthful* if it truthfully indicates the option that is advantageous to the Receiver. After receiving the message chosen by the Sender, the Receiver chooses an option, which then determines payments. Both players are fully informed about the rules of the game, but Receivers are never informed about the specific monetary consequences of either of the two options.

---

[5] Effects of social information have been studied in a number of other contexts. In ultimatum games, Knez and Camerer (1995) and Bohnet and Zeckhauser (2004) examine effects of information about other proposer offers on proposer and responder behavior, finding evidence of a pre-existing norm of equity. Several authors study the role of social information in achieving social learning and conditional cooperation in coordination games (Berg, Dickhaut and McCabe, 1995; Fischbacher, Gachter and Fehr, 2001; Schotter and Sopher, 2003; Chaudhuri, Graziano, and Maitra, 2006; Eckel and Wilson, 2007); in this work, unlike ours and like studies by Chen, et. al (2009) on on-line participation in MovieLens and Duffy and Feltovich (1999) on learning in ultimatum and best shot games, the social information is potentially payoff relevant. Recent field experiments on charitable contributions document that subjects contribute more often when they believe a higher fraction of their peers contribute (Frey and Meier, 2004) and contribute more when told that a prior contributor contributed more (Shang and Croson, 2008); in this work, the social information can also be payoff relevant by signaling the virtue of the charity. Overall, this work provides evidence of contagion in different contexts, but with potential channels of effect that are not at play in our experiments. Lindbeck, Nyberg and Weibull (1998) also model contagious preferences, in the context of stigma for welfare; however, they *assume* that the stigma from welfare is negatively related to the proportion of a relevant peer group on welfare and study implications of this assumption. In contrast, we are interested in testing for the presence of contagion in honesty.

In our experiment, we focus on a single set of payment options (while randomly varying the A/B labels attached to the two options). In one, the Sender receives $6 and the Receiver obtains $3, while in the other, the Sender receives $4 and the Receiver obtains $6.[6]

Our objective is to study how different perceptions of the truthfulness of other Senders affects Sender behavior. To do this, we use a between-subjects design where we expose different groups of Senders to different treatments designed to alter perceptions of other Sender behavior. In the control treatment, given to an initial session of subjects, no information on other Sender behavior is given. Using outcomes from the control treatment, Senders in subsequent sessions are given information in the form of the following statement:

> "Out of 20 Sender messages from past sessions of this experiment, with
> identical payment options, X (=Y%) were UNTRUTHFUL and (20-X)
> (=(100-Y)%) were TRUTHFUL."

Four treatments of this form are considered: Y=15% (heavily truthful), Y=40%, Y=60%, and Y=85% (heavily untruthful). In all treatments, the higher percentage is reported first (so that, for example, when Y=40%, the number and percent of truthful messages from past sessions is indicated first). Our approach is similar to that used in other experimental papers in the social influence literature. Frey and Meier (2004), for example, report two different percentages of past students who contribute to a charity based on different outcomes from a recent semester and, alternatively, a ten year interval. Bicchieri and Xiao (2009) report different shares of "fair" choices (40% and 60%) from a past session and argue in their paper that the information is truthful because they can define a past session to satisfy either claimed percentage. We designed our statements to highlight the selection of a subset of Sender messages

---

[6] There is no obvious choice of payment options for our experiment. We conducted preliminary surveys on alternative options that varied (1) the gain to the Sender from lying G_S (assuming Receiver acceptance of recommendations), and (2) the corresponding Receiver loss L_R. Consistent with expectations, incentives to lie rise with G_S and fall with L_R. Our survey evidence implied an approximate Sender propensity for truthfulness equal to 58 percent for $G_S=2$ and $L_R=3$ (our chosen options). Armed with this evidence – and the conjecture (wrong as it turned out) that actual dollar stakes would raise incentives to lie – we settled on the indicated options.

and were careful not to state or imply that the reported messages represent a general pattern.[7]

We use these treatments to test the following:

Contagion Hypothesis. The likelihood of untruthful behavior by a Sender rises with the perceived likelihood that other Senders are untruthful.

In our experiment, we are implicitly jointly testing (1) whether the treatments are believed, and (2) effects of treatment-induced perceptions of other Sender behavior. Hence, if we find no significant effect of a treatment, we cannot reject the "contagion hypothesis" per se. However, if we find a significant effect – in the predicted direction – we can reject the null hypothesis of no contagion.

In all treatments, we provide Senders with general information on the propensity of Receivers to accept their recommendations. Based on results from Gneezy's (2005) experiments (where 78 percent of Receivers followed the Sender recommendations), we tell all Senders the following:

"In past experiments like this one, roughly 8 out of 10 Receivers chose the Option recommended by their Senders."

Receivers are not given this information, and Senders are so informed. The instructions given to Receivers and Senders (for the heavily untruthful treatment), respectively, are provided in the Appendix.

To verify that Senders generally believe that Receivers would accept their recommendations, we follow Gneezy's (2005) approach, asking them to predict their Receiver's choice and paying them for a correct prediction. Overall, 73.4% of Senders predicted that their Receiver would accept their recommendation.[8] These

---

[7] A norm in experimental economics is that the experimenter be honest with his/her subjects. We obey this norm with our approach. However, our treatments are intended to influence perceptions. We note that experimental designs with such objectives are common in the experimental economics literature. Prominent examples are influential papers that report a subject's "awarded" status "to suggest to the (other) subjects that the high status was deserved" when in fact it was randomly assigned (Ball, et al., 2001), that expose subjects to resume's with fictitious racial profiles (Bertrand and Mullainathan, 2004), that elicit contributions for a public project given fictitious variation in seed money (List and Reiley, 2002), and that use a standard experimental protocol to not inform subjects that they will be playing in subsequent rounds or roles (e.g., Binmore, et. al, 1985; Duffy and Kornienko, 2009). See Bonetti (1998) for a lucid discussion of this topic.

[8] In principle, risk aversion could motivate an "accept" prediction by truthful Senders and a "reject" prediction by untruthful Senders. However, the proportion of truthful Senders predicting Receiver

results indicate that Senders generally expect their recommendations to be followed; hence, their choices reflect a concern for the "fairness" / morality of lying, and not strategic motives. As it turned out, 73 percent of our Receivers followed their Sender recommendations.

The experiment was conducted in undergraduate economics classes at the University of Arizona in Spring, 2008 and Spring, 2009. In total, there were 233 Sender/Receiver pairs. Receivers were in different classes than any of the Senders. Anonymity of all participants was ensured by identifying subjects with a randomly assigned identification number that was also used to match Senders to Receivers. Class Rosters were used to ensure that no student participated more than once.[9] The experiment took approximately 10 minutes to run. Subject participation was purely voluntary. Subjects were so informed and instructed to communicate only with the experimenter during the experiment and were carefully monitored to this end. Control treatments were run in each Sender class to control for any potential individual class/course effects.

Table 1 reports the number of Senders exposed to each of the different treatments, and summarizes our results. Table 2 reports results of a probit regression of truthful (y=1) vs. untruthful (y=0) choices on the treatments and individual class fixed effects. Relative to the control, one treatment has a significant impact on subjects' propensity to be truthful. In the "heavily untruthful" treatment (Y=85%), the proportion of untruthful messages rises from 41 percent (in the control) to approximately 81 percent (under the treatment), almost doubling. Other treatments have no significant effect. *We thus find support for the contagion hypothesis in the sense that a strong propensity for untruthfulness is contagious.*

In principle, two other effects discussed in the literature might also be at play in our experiment. First is the effect our treatments might have in creating a focal point (see, for example, Crawford, et al., 2008). Although focal points generally serve as coordination mechanisms that are not relevant in our simple experiment, the

accept decisions (74.8%) is essentially identical to the proportion of untruthful Senders predicting accept decisions (72.0%) in our experiment.

[9] There was no overlap between the Receiver class and any of the Sender classes. Two students who were enrolled in two of the Sender classes were not present when the second class experiment was performed.

mention of others' behavior may have a focusing effect (Krupka and Weber, 2009). However, any focusing effect would arise in any of our treatments. For example, both Y=60% and Y=85% treatments could focus subjects on the untruthful message. The absence of a significant impact of the Y=60% treatment (vs. the control) and the presence of a significant impact of the Y=85% treatment (vs. the Y=60% treatment) argues against "focusing" as an explanation for our main result. A number of scholars have suggested that subjects mechanistically respond to reference (or anchoring) points (see Shang and Croson, 2008, for discussion), implying in our case that stronger treatments will be more effective in stimulating compliant behavior. Although the lack of monotonic effect in our experiments loosely argues against a monotonic "mechanistic" response, we believe the distinction between these two explanations – cognitive reference point vs. social influence – is empty. Indeed, as argued below, we believe our results are likely to be explained by hard-wired contagious preferences that spur a mechanistic response to the social information that we provide.

Second – and closely related – there is the potential for experimenter demand effects, with subjects trying to do what the experimenter appears to want them to do (see Duffy and Kornienko, 2009, for an excellent discussion). Belying such effects in our experiments is again the absence of significant impact of our other (Y=15%, 40%, 60%) treatments which, were there an experimenter demand effect, would also influence behavior.

Finally, Sutter (2009) documents the importance of "sophisticated deception," when a Sender tells the truth anticipating that his recommendation will not be followed. Sutter (2009) proposes a different measure of deception that includes sophisticated deceivers and excludes "benevolent liars" who lie anticipating that their Receiver will reject their recommendation. In our experiment, the treatments concern raw actions of other subjects (truthful vs. untruthful); hence, we are also principally interested in effects on raw actions (as reported above), rather than on Sutter's measure of sophisticated deception. The conclusions of Table 1 nonetheless extend to Sutter's measure, although the impact of our treatment on "sophisticated deception" is attenuated; the proportion of Sutter-truth-tellers is 55.7 percent under

the control and 30.8 percent under the heavily untruthful (Y=85%) treatment, a
difference that is statistically significant (z=3.06).[10]

B. Arizona Survey Evidence

Like Gneezy (2005), we also obtained survey evidence on our research
hypothesis. In a survey of 174 economics undergraduates in the Fall of 2007, we
asked their response to a hypothetical situation in which they could be untruthful (to
their material advantage) or truthful (to their material disadvantage). The specific
situation posed is as follows:

> "Suppose that you have been visiting a country called Bayeb. Before leaving
> the country permanently, you must sell your used car. A local person
> (unknown to you) agrees to buy the car for US $2,000 and pay you in cash.
> However, you know that the radiator in your car is not functioning properly
> and the problem will only become noticeable after 2 months. The buyer does
> not know about the problem. If you tell him/her about the problem, then you
> have to reduce the price of the car by US $250 and sell it for US $1,750.
> However, if you do not reveal the problem, then you can sell the car for US
> $2,000 and the buyer will have to fix the car after 2 months, spending US
> $250. Would you tell the buyer about the radiator problem?"

Because the buyer is not known to the individual, and the individual is leaving
the country permanently, before the problem can be discovered, anonymity is
assumed and social / institutional sanctions are impossible. We consider three
treatments: (1) A control with no further information. (2) A "truthful" treatment in
which the respondent is told the following:

> "Surveys in Bayeb indicate that, in a situation like yours, 9 out of 10 people
> would tell the buyer about the radiator problem."

(3) An "untruthful" treatment in which the respondent is told the following:

---

[10] The proportions of Sutter-truth-tellers are 56%, 57.7% and 60.6% under the other (Y=15%, 40% and
60%) treatments. Note that Sutter (2009) finds almost no "benevolent liars" in his experiment. In
contrast, the proportion of "benevolent liars" in our subject pool (14.2%) is roughly the same as the
proportion of "sophisticated liars" (12.4%); we also find no clear pattern in this behavior across
treatments (for example, in the proportion of liars who are benevolent or the proportion of truth-tellers
who are sophisticated liars). These observations loosely suggest that the predictions of our Reject-
predicting subjects may be random, reflecting an anticipation that the Receiver essentially flips a coin
when making his choice.

"Surveys in Bayeb indicate that, in a situation like yours, 9 out of 10 people would not tell the buyer about the radiator problem."

Table 3 reports survey results. Relative to the control, the truthful treatment increases the proportion of truthful respondents by ten percent; conversely, the untruthful treatment increases the proportion of untruthful respondents by almost twenty percent. Only the second effect is statistically significant. Hence, we again have evidence of contagion in the sense that a perceived propensity for others to be highly untruthful is contagious.

C. The India Laboratory Experiment

In our Arizona experiment, we find that information indicating a strong peer propensity for dishonesty promotes untruthful behavior. A central motive for our work is to study the potential for contagion in the other direction: In a country where corruption is high, and a propensity for dishonesty correspondingly high, can subjects be spurred to more truthful conduct by information suggesting a norm of honesty? India – with a corruption index in the highest tier of countries – is arguably an ideal country in which to examine this question.

In the Spring of 2009, we conducted a deception experiment with a set of 60 Sender/Receiver pairs of undergraduate students at Jadavpur University in Calcutta. Like most university experimental labs, Jadavpur maintains a roster of willing experimental participants and regularly announces opportunities for participation in experiments. Our announcements were made in English, and were only made in Departments where English fluency is required (most of the University, excepting the Bengali major). In the India experiment, the following two payoff options were posed:

Option A: 160 Rupees to you (the Sender) and 160 Rupees to the other student (the Receiver).

Option B: 200 Rupees to you (the Sender) and 100 Rupees to the other student (the Receiver).

As in Arizona, option labels were varied randomly. The payoffs were designed to (a) have the same ratio of Receiver loss to deceit and Sender gain (3/2) as in our Arizona experiment, (b) to meet minimum payment requirements, and (c) to give substantial

stakes to the choices made. Although 40 Rupees (the Sender gain from dishonesty and Receiver acceptance) are less than one U.S. dollar, average daily per capita consumption expenditures in India are less than 19 Rupees in rural areas and 35 Rupees in urban areas.[11] Put differently (quoting Fehr, Hoff and Kshetramade, 2008), "Fifty rupees are roughly equal to a day's skilled wage." The stakes in our experiment can therefore be considered substantial in context.

We conducted two treatments, a control with no information about Sender behavior in prior sessions of related experiments and a strongly truthful treatment in which Senders were given the following information:

"Out of 15 Sender messages from a past session of this experiment here in Calcutta, 13 out of 15 (85%) were TRUTHFUL and 2 out of 15 (15%) were UNTRUTHFUL."

Our initial control treatment responses in Calcutta gave us the 15 Sender messages satisfying this statement.[12]

Table 4 reports results from the Calcutta experiment. We find that the honest treatment leads to a significantly higher proportion of truthful messages than in the control, although the level of significance (p=.067) is greater than five percent (two-sided). Under the treatment, the proportion of honest messages is more than fifty percent higher than under the control, 67.7% vs. 44.8%. The proportion of Senders predicting Receiver acceptance is high (78.3%) and the proportion of Receivers accepting their Sender recommendations is also high (70.2%), although less than in Gneezy's (2005) experiments and slightly less than in Arizona.[13]

3. Explanations for Contagion

---

[11] See "Household Consumption Expenditure in India (January-June 2004)," NSSO, Government of India, 23 November, 2005.

[12] As in Arizona, all Senders were told that roughly 80% of Receivers accepted their Sender recommendations in a similar prior experiment; none of this information was provided to Receivers and Senders were so informed.

[13] In the India experiment, a slightly higher fraction of truthful Senders predict Receiver acceptance than do untruthful Senders (82.3% vs. 73.1%), but the difference is not statistically significant (z=.852). Likewise, as indicated in Table 4, a slightly higher fraction of control subjects predict Receiver acceptance, but again the difference is not statistically significant (z=.806).

Our results indicate that some of our subjects have an aversion to lying that increases with the perceived propensity for honesty in a relevant peer group. Why might this be true? Because our experimental design ensures that there is no scope for social sanctions or building social esteem, and the information we provide is irrelevant to experimental payoffs for Sender and Receiver, the contagion we observe does not reflect standard theoretical motives for "conformity" (see the introduction). This leaves two alternative explanations: First, perhaps subjects have social (other-regarding) preferences. If so, then information about other Senders' behavior can potentially be relevant to the utility a subject derives from different actions (message choices). Second, alternatively, contagion may be "hard-wired": Subjects may have a built-in contagion trait that, as a reflex, prompts them to change their aversion to lying in response to what other people do (copying the majority).

A. Social Preferences

Two theories of social preferences (suitably modified) are consistent with Gneezy's (2005) findings on deception behavior and are therefore the most natural candidates for explaining our results: 1) the relative payoff preferences of Fehr and Schmidt (1999) or Bolton and Ockenfels (2000), and 2) the guilt aversion posited by Charness and Dufwenberg (CD, 2006) and Battigalli and Dufwenberg (BD, 2007).

With Fehr-Schmidt preferences, agents are averse to inequality, whether due to obtaining a higher payoff than others or, worse, a lower payoff. Alternately, if they are "spiteful" (see Levine, 1998; or Fehr, Hoff and Kshetramade, 2008), they may benefit from a higher relative payoff. Such social preferences alone do not imply an effect of information about other Senders' propensity for honesty on a Sender's utility-maximizing decision. Necessary for such an effect is that a Sender's reference group – the group of subjects to whom a Sender compares himself – be a broader population than the Receiver who is directly affected by the Sender's decision. This property is controversial; for example, Ellingsen and Johannesson (2008, p. 994) express skepticism that agents care about the outcomes from others' actions in choosing their own conduct. Even under a "broad reference group" premise (the Sender compares himself to all subjects in the experiment), we can show that generalized Fehr-Schmidt preferences, modified to be consistent with Gneezy's

(2005) findings, do not imply contagion.[14] Indeed, for our Arizona experiment, they imply the opposite: incentives for honesty *fall* with the perceived propensity for honesty in the reference population of Senders.[15]

Perhaps guilt aversion offers more promise. CD and BD posit that subjects are averse to disappointing their partners: If a Receiver obtains a payoff that is less than he or she expects to obtain (where the "Receiver expectation" is based on the Sender's belief about the Receiver's beliefs), then the Sender suffers a guilt aversion penalty that is proportional to the extent of the shortcoming. This logic, we believe, is likely to be important in explaining subjects' behavior in deception experiments (see Charness and Dufwenberg, 2005). The question here is whether it can explain the *contagion* that we observe.

In principle, the answer is "yes" if our treatments affect a Sender's beliefs about the Receiver's expectations. Suppose that a higher Sender expectation about the propensity for other Senders to be untruthful (as induced by our "heavily untruthful" treatment) prompts Senders to believe that Receivers also believe that there is a higher Sender propensity for untruthfulness. Then, given an assumed mechanical acceptance of Sender recommendations, the Sender expects the Receiver to expect a lower payoff, which lowers the guilt aversion penalty to lying and thus prompts more Senders to lie.

Although aspects of our experimental design mitigate such effects,[16] we nonetheless test for them directly. We do so by measuring Sender beliefs about

---

[14] Gneezy (2005) rightly points out that pure Fehr-Schmidt preferences predict that Sender incentives for dishonesty rise with the Receiver's high payoff, contrary to his experimental results. The addition of a social welfare component to preferences, and a utility penalty to deceit, cures this inconsistency.

[15] Details are available in our expanded paper. This conclusion rests on the plausible premises that (1) the reference group is the overall population of experimental subjects and (2) the inequity disutility functions are weakly convex, implying that larger inequities are not better, per unit, than smaller ones. The second premise mimics Bolton and Ockenfels (2000) Assumption 3. Intuitively, a higher propensity for Sender honesty increases the probability that Receivers obtain their high payoff and that other Senders obtain their low payoff. This raises the net benefit to dishonesty by lowering the implied cost of inequity with respect to Receivers and lowers the net benefit to dishonesty by raising the cost of inequity with respect to other Senders. In our Arizona experiment, the former (Receiver) effect dominates the latter (Sender) effect because disadvantaged Receivers obtain less than disadvantaged Senders; hence incentives are tilted toward dishonesty, contrary to the contagion hypothesis. Similar logic applies to both spiteful (Fehr, et al., 2008) and Bolton-Ockensfels (2000) preferences.

[16] Receivers in our experiment are never told the payoffs available in the game and, hence, have no basis for disappointment. Senders are told this and also know that the information about Sender behavior in prior sessions – our treatment – is not provided to Receivers. Senders are also told that

Receiver beliefs (about Senders' propensity to lie) and evaluating their impact on Senders' decisions on whether or not to lie. We stress that this exercise is NOT a test of guilt aversion per se; guilt aversion does not predict that the Sender beliefs we measure will necessarily alter Sender decisions to lie or not.[17] However, in order for guilt aversion to fully explain the *contagion* that we observe, our treatments must not have an impact on Sender message decisions that is distinct from the impact of Sender beliefs. That is, our null hypothesis – the guilt aversion explanation for contagion – is that our treatment effects are zero once we account for Sender beliefs about Receiver beliefs.

To measure Sender beliefs about Receiver beliefs, we asked Receivers the following in our India experiment and a subset of our Arizona experiments (28 control treatment Sender / Receiver pairs and 29 "heavily untruthful" treatment subject pairs):

> "We now ask you to predict the proportion of Senders in this experiment that sent truthful messages… If your prediction is correct (within five percentage points of the actual proportion, plus or minus), you will receive an addition $1 (20 Rupee) payment."

Receivers were then asked to circle one of twenty five-percentage-point bands (from 0-5% to 96-100%). The question was posed after Receivers made their option choice. Similarly, Senders were asked the following (after they made their message choice):

> "Your Receiver will indicate to us his/her belief about the proportion of Senders that are truthful. After selecting the payment option and before receiving payment from the experiment, your Receiver will indicate that out of 100 Senders, he/she believes that X percent are truthful. We now ask you

---

Receivers generally accept their recommendations mechanically, and we have evidence that this statement is believed in all treatments (Tables 1 and 4). If Senders believe that Receivers internalize the treatment information delivered only to Senders, they should also expect Receivers to revise their decisions on whether or not to accept or reject Sender recommendations.

[17] We establish this formally in our expanded paper. Intuitively, there can be two effects of Sender beliefs. The first is the pro-contagion effect described in the text (for a given Sender belief about the probability of Receiver acceptance). The second is due to a (rational) Sender belief that, with a higher Receiver assessment of the probability of Sender truthfulness, the Receiver accepts the Sender's recommendation with higher probability; this raises the Sender's (self-interested) incentive to lie, countering the first (contagion) effect. Either effect can dominate, implying no clear prediction from guilt aversion theory about the impact of Sender beliefs on Sender deception decisions.

to predict your Receiver's indicated belief (X). If your prediction is correct (within five percentage points of the actual choice, plus or minus), you will receive an additional $2 (20 Rupee) payment."

In the Arizona experiment, average Sender beliefs about Receiver beliefs (about the proportion of truthful Senders) were 58.6% under the control (using midpoint values) and 38.0% under the heavily untruthful (Y=85%) treatment, a significant difference (z=3.264). In the India experiment, in contrast, average Sender beliefs were 65.1% under the control and 69.1% under the heavily truthful (X=15%) treatment, an insignificant difference (z=.85).[18]

Given the surprising correlation between our treatments and Arizona Sender beliefs, Table 5 reports probit estimations to test for distinct effects of Sender beliefs and our treatments on Sender message choices (virtually identical results are obtained with logit). The Sender beliefs are statistically insignificant and, more importantly, do not confound the distinct effects of our treatments.[19] Hence, if guilt aversion underpins subjects' aversion to lying in our experiments – and nothing we have done suggests otherwise – then our treatments change the guilt aversion parameters in subjects' preferences, a contagion effect that is not itself explained by the theory.

B. Hard-wired Contagion

Contagion in honesty may be a reflexive response of subjects – a hard-wired trait that tells them to "do as others do" in these types of situations. This explanation for our findings is promising, we believe, but also shallow. It begs the deeper question: Why would such a trait evolve? That is, why is a contagion trait advantageous from an evolutionary perspective?

Two observations form the basis for a more complete inquiry into this subject (Innes, 2009), drawing broadly on the evolutionary strategy literature (e.g., Frank,

---

[18]Average Receiver beliefs were virtually identical in the two experiments, 55.8% in Arizona and 54.4% in India.

[19]In principle, subjects' expectations of Receiver beliefs could depend upon their decisions, implying that they are endogenous. Unfortunately, in our experiments, we have no exogenous instruments, distinct from our treatments, with which to identify Sender beliefs. However, the only conceivable mechanisms for endogeneity (that we can envision) – such as risk aversion or subjects projecting expectations based on their own behavior – imply a positive relationship between truth-telling and Sender beliefs that would bias our estimations against the treatment effects that we find in Table 5. Our estimations thus provide evidence in favor of distinct treatment effects that is robust to potential endogeneity.

1987; Bergstrom, 2002; Guth and Kliemt, 1994; Sobel, 2005). First, there can be network effects that motivate conformity with the conduct of others (Katz and Shapiro, 1986; Banerjee and Besley, 1990). Although our experiment contains no network effects, the posed situation of moral dilemma is arguably sufficiently similar to others in which network effects are present that a contagion trait, motivated by the latter situations, kicks in. In particular, honesty can be central to advantageous outcomes in games of cooperation. In such games, honesty can be advantageous when most others are honest because honest people only engage in profitable partnerships with other honest people. Likewise, dishonesty can be advantageous when most others are dishonest because honest people are exploited in joint ventures and therefore withdraw from them. Second, however, this logic only motivates conformist equilibria in which all are honest or all are dishonest. A contagion trait – telling an individual to change preferences in response to social cues – can be advantageous when there is trade between groups that have evolved to different equilibria. Only those who are "contagious" will be able to partner with others in a different group in which a different norm predominates.

4. Conclusion

We find evidence that honesty can be contagious when subjects are otherwise predominantly dishonest (Calcutta) and dishonesty can be contagious when subjects are otherwise predominantly honest (Arizona). These responses shed some light on population dynamics in truthfulness and corruption that may help to explain societal tendencies to be in one camp or the other, highly honest or highly dishonest. Normatively, they suggest value to a culture of honesty in an organization by indicating the fragility of truthful behavior; even with small stakes, our Arizona subjects flocked to the dishonest course when primed with a social pass-go to do so. Conversely, they suggest promise for countering corrupt impulses in the developing world if perceptions of norms can be reversed.

Of course this begs the question: By what mechanism can norms be changed? Recent findings suggest that this may be tough. Rode (2008), for example, finds that dishonesty is insensitive to cooperative priming, and Fisman and Miguel (2007) find

that foreign diplomats do not respond to American values of lawful behavior. However, empirical evidence indicates that aid and trade reduce corruption (Tavares, 2003; Gokcekus and Knorich, 2006; Innes and Mitra, 2009). Our results suggest a coarse mechanism for this effect, but leave much unanswered. For example, what determines whether "honest norm" partners bend to the norms of "dishonest norm" partners or vice versa? Our Arizona survey results reveal a sensitivity to local norms, but more work is needed to determine how a subject's exposure to another country's norms effects his behavior in his own country. Moreover, if (as we suggest in Section 3) contagion in honesty is hard-wired and motivated by economic interactions between societies with different norms, then contagion will be stronger in groups that trade more with outside groups, which in turn implies links between the extent of trade and the responsiveness of local norms to trade relations. For example, if an "honest norm" agent seeks to trade in a "dishonest norm" country that is relatively closed, then the honest trader is likely to bend to local (dishonest) norms, rather than vice versa. These questions and conjectures lend themselves to further experimental work that can illuminate not only the nature of the contagion we identify, but also how it can be exploited for positive social ends and what implications it has for one of the key pillars of the globalization debate: benefits of trade in reducing corruption.

References

Acemoglu, D., S. Johnson, and J. Robinson. 2001. "The Colonial Origins or Comparative Development: An Empirical Investigation." *American Economic Review,* 91: 1369-1401.

Akerlof, G. 1980. "A Theory of Social Custom, of Which Unemployment May Be One Consequence." *Quarterly Journal of Economics* 94: 749-75.

Ball, S., C. Eckel, P. Grossman, and W. Zame. 2001. "Status in Markets." *Quarterly Journal of Economics* 116: 797-817.

Banerjee, A. "A Simple Model of Herd Behavior." 1992. *Quarterly Journal of Economics* 107: 797-817.

Banerjee, A., and T. Besley. 1990. "Peer Group Externalities and Learning Incentives: A Theory of Nerd Behavior." John M. Olin Discussion Paper no. 68. Princeton, N.J.: Princeton University.

Bardsley, N. 2008. "Dictator Game Giving: Altruism or Artefact?" *Experimental Economics* 11: 122-33.

Battigalli, P. and M. Dufwenberg. "Guilt in Games." 2007. *American Economic Review* 97: 170-6.

Berg, J., J. Dickhaut, and K. McCabe. 1995. "Trust, Reciprocity and Social History." *Games and Economic Behavior* 10: 122-42.

Bergstrom, T. 2002. "Evolution of Social Behavior: Individual and Group Selection." *Journal of Economic Perspectives* 16: 67-88.

Bernheim, B.D.. "A Theory of Conformity." 1994. *Journal of Political Economy* 102: 841-77.

Bertrand, M., and S. Mullainathan. 2004. "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94: 991-1013.

Bicchieri, C., and E. Xiao. 2009. "Do the Right Thing: But Only if Others Do So*." Journal of Behavioral Decision Making* 22: 191-208.

Bikhchandaria, S., D. Hershleifer, and I. Welch. 1992. "A Theory of Fads, Fashion, Custom, and Cultural change as Informational Cascades." *Journal of Political Economy* 100: 992-1026.

Binmore, K., A. Shaked, and J. Sutton. 1985. "Testing Noncooperative Bargaining Theory: A Preliminary Study." *American Economic Review* 75: 1178-80.

Bohnet, I. and R. Zeckhauser. 2004. "Social Comparisons in Ultimatum Bargaining." *Scandinavian Journal of Economics* 106: 495-510.

Bolton, G. and A. Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review* 90: 166-93.

Bonetti, S. 1998. "Experimental Economics and Deception." *Journal of Economic Psychology* 19: 377-95.

Cason, T. and V. Mui. 1998. "Social Influence in the Sequential Dictator Game." *Journal of Mathematical Psychology* 42: 248-65.

Charness, G. and M. Dufwenberg. 2006. "Promises and Partnership." *Econometrica* 74: 1579-1601.

Charness, G. and M. Dufwenberg. 2005. "Deception: The Role of Guilt." Working Paper, UC Santa Barbara and University of Arizona.

Chaudhuri, A., S. Graziano and P. Maitra. 2006. "Social Learning and Norms ina Public Goods Experiment with Inter-Generational Advice." Review of Economic Studies 73: 357-80.

Chen, Y., M. Harper, J. Konstan and S. Xin Li. 2009. "Social Comparisons and Contributions to Online Communities: A Field Experiment on MovieLens." *American Economic Review,* Forthcoming.

Cialdini, R. and N. Goldstein. 2004. "Social Influence: Compliance and Conformity." *Annual Review of Psychology* 55: 59-621.

Crawford, V., U. Gneezy and Y. Rottenstreich. 2008. "The Power of Focal Points Is Limited: Even Minute Payoff Asymmetry May Yield Large Coordination Failures." *American Economic Review* 98: 1443-58.

Duffy, J. and N. Feltovich. 1999. "Does Observation of Others Affect Learning in Strategic Environments? An Experimental Study." *International Journal of Game Theory* 28: 131-52.

Duffy, J. and T. Kornienko. 2009. "Does Competition Affect Giving?" Working Paper, University of Pittsburgh and University of Edinburgh.

Eckel, C. and R. Wilson. 2007. "Social Learning in Coordination Games: Does Status Matter?" *Experimental Economics* 10: 317-29.

Ederer, F. and E. Fehr. 2007. "Deception and Incentives: How Dishonesty Undermines Effort Provision." Discussion Paper No. 3200, IZA, Bonn.

Ellingsen, T., and M. Johannesson. 2008. "Pride and Prejudice: The Human Side of Incentive Theory." *American Economic Review* 98: 990-1008.

Fehr, E., and U. Fischbacher. 2004. "Social Norms and Human Cooperation." *Trends in Cognitive Sciences* 8: 185-90.

Fehr, E., K. Hoff and M Kshetramade. 2008. "Spite and Development." *American Economic Review* 98: 494-99.

Fehr, E. and K. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics* 114: 817-68.

Fischbacher,U., S. Gachter and E. Fehr. 2001. "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment." *Economics Letters* 71: 397-404.

Fisman, R., and E. Miguel. 2007. "Corruption, Norms and Legal Enforcement: Evidence From Diplomatic Parking Tickets." *Journal of Political Economy*, 115: 1020-48.

Frank, R. 1987. "If Homo Economicus Could Choose His Own Utility Function, Would He Want One with a Conscience?" *American Economic Review* 77: 593-604.

Frey, B. and S. Meier. 2004. "Social Comparisons and Pro-Social Behavior: Testing 'Conditional Cooperation' in a Field Experiment." American Economic Review 94: 1717-22.

Gneezy, U. 2005. "Deception: The Role of Consequences." *American Economic Review* 95: 384-94.

Gokcekus, O. and J. Knorich. 2006. "Does Quality of Openness Affect Corruption?" *Economics Letters* 91; 190-6.

Guiso, L., P. Sapienza, and L. Zingales. 2004. "The Role of Social Capital in Financial Development." *American Economic Review* 94: 526-56.

Guth, W. and H. Kliemt. 1994. "Competition or Cooperation: On the Evolutionary Economics of Trust, Exploitation and Moral Attitudes." *Metroeconomica* 45: 155-87.

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

Innes, R. "A Theory of Moral Contagion." 2009. Working Paper, U.C. Merced.

Innes, R. and A. Mitra. 2009. "Trade and Corruption." Working Paper, University of Arizona.

Katz, M., and C. Shapiro. 1986. "Technology Adoption in the Presence of Network Externalities." *Journal of Political Economy* 94: 822-41.

Knez, M. and C. Camerer. 1995. "Outside Options and Social Comparison in Three-Player Ultimatum Game Experiments." *Games and Economic Behavior* 10: 65-94.

Krupka, E. and R. Weber. 2009. "The Focusing and Informational Effects of Norms on Pro-Social Behavior." *Journal of Economic Psychology,* Forthcoming.

Levine, D. 1998. "Modeling Altruism and Spitefulness in Experiments." *Review of Economic Dynamics* 1: 593-622.

Lindbeck, A., S. Nyberg, and J. Weibull. 1999. "Social Norms and Economic Incentives in the Welfare State." *Quarterly Journal of Economics* 114: 1-35.

List, J. "On the Interpretation of Giving in Dictator Games." 2007. *Journal of Political Economy* 115: 482-93.

List, J., and D. Reiley. 2002. "The Effects of Seed Money and Refunds on Charitable Giving: Experimental Evidence from a University Capital Campaign." *Journal of Political Economy* 110: 215-33.

Rode, J. 2008. "Truth and Trust in Communication – Experiments on the Effect of a Competitive Context." Working Paper No. 0804, Universitat Mannheim.

Sanchez-Pages, S. and M. Vorsatz. 2006. "An Experimental Study of Truth-Telling in a Sender-Receiver Game." Working Paper, University of Edinburgh and Maastrict University.

Schotter, A. and B. Sopher. 2003. "Social Learning and Coordination Conventions in Intergenerational Games: An Experimental Study." *Journal of Political Economy* 111: 498-529.

Shang, J. and R. Croson. 2008. "Field Experiments in Charitable Contribution: The Impact of Social Influence on the Voluntary Provision of Public Goods." Working Paper, University of Pennsylvania.

Sobel, J. 2005. "Interdependent Preferences and Reciprocity." *Journal of Economic Literature* 43: 393-436.

Sutter, M. 2009. "Deception through Telling the Truth? Experimental Evidence from Individuals and Teams." *Economic Journal* 119: 47-60.

Tavares, J. 2003. "Does Foreign Aid Corrupt?" *Economics Letters* 79: 99-106.

Zak, P. and S. Knack. 2001. "Trust and Growth." *Economic Journal* 111: 295-321.

Instructions for Receivers   **Your Registration Number:** _____

Introduction: Your registration number is written at the top of this page. This number will be used to identify you for payment. From now on, you should not communicate in any way with the other participants until the end of the session. If you have any question at any time, please raise your hand and one of us will help you.

The Experiment: This is a short experiment in decision making. In this experiment,

• You have been randomly matched with another student from a different class, who we will call your "Sender."

• Neither of you will ever know the identity of the other.

• The money that you earn will be paid to you after you have completed this questionnaire; payment will be private (in an envelope) and in cash.

• Two possible monetary payment options (A and B) are available to you and your Sender in the experiment. We showed the two payment options to your Sender.

• **YOU** will choose **ONE** of the two options, which will determine the payments to the two of you. The only information you will have is the message your Sender sends to you.

• Two possible messages could be sent:

    Message A: "Option A will earn you more money than Option B."
    Message B: "Option B will earn you more money than Option A."

• Your Sender decided to send you Message _____

• We now ask you to choose either Option A or Option B. Your choice will determine the payments in the experiment. You will never be told what sums were actually offered in the option not chosen (that is, if the message sent by your Sender was true or not). Moreover, you will never be told the sum your Sender actually receives.

We now ask you to choose ONE of the two options. **(Please circle one)**

        Option A   **Option B**

Your gender (please circle one):   **Male**   **Female**

Instructions for Senders   **Your Registration Number:** _____

Introduction: Your registration number is written at the top of this page, and on the attached ribbon. This number will be used to identify you. Please tear off and keep the attached ribbon, and bring it with you to class next week for payment. We will never know the real identity of any participant in this experiment.

The Experiment: This is a short experiment in decision making.

• You will be randomly matched with another student from a different class, who we will call your "Receiver."
• Neither of you will ever know the identity of the other.
• The money that you earn will be paid to you during the last class session next week, in cash and in confidence.
• Two possible monetary payments are available to you and your Receiver in the experiment. The two payment Options are:

  Option A: $6 to you and $3 to the other student (your Receiver).
  Option B: $4 to you and $6 to the other student (your Receiver).

• You know what the Options are, but your Receiver **DOES NOT**.
• You will choose between **ONE** of two possible messages, which will be sent to your Receiver:

  Message A: "Option A will earn you (the Receiver) more money than Option B."
  Message B: "Option B will earn you (the Receiver) more money than Option A."

• Based only on your message, your Receiver will **CHOOSE ONE** of the two Options. Your Receiver's choice will determine the payments in the experiment. Receivers will never know the true Options, or the sums to be paid to you under the different Options.

• **Information from Past Sessions**:

  • Out of **20 Sender messages** from past sessions of this experiment, with identical payment options,
            17 out of 20 (85%) were UNTRUTHFUL, and
            3 out of 20 (15%) were TRUTHFUL.

  (A "TRUTHFUL" message indicates the Option that actually earns the Receiver more money.)

  • In past experiments like this one, roughly **8 out of 10 Receivers chose the Option recommended by their Senders**.

  • Your Receiver does not have this information.

** I choose to send the following message (please circle ONE): **

              <u>**Message A**</u>   <u>**Message B**</u>
              (Sender Instructions continued)

To satisfy our curiosity, could you please answer the following question:

Which option do you think your Receiver will choose based upon the message you sent? If you correctly predict your Receiver's choice, we will pay you an additional $1. Please circle ONE.**

I believe my Receiver will choose   <u>Option A</u>

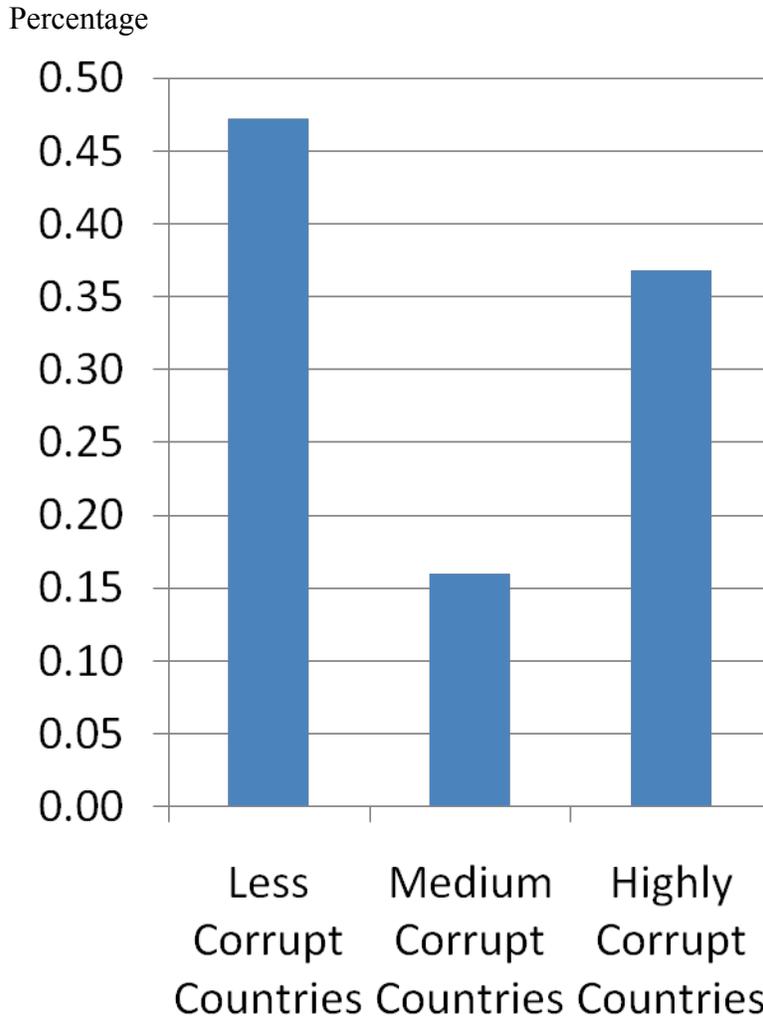**I believe my Receiver will choose**   <u>Option B</u>

Your gender (please circle one):   <u>Male</u>   <u>Female</u>

Figure 1A. Distribution of World GDP by Level of Corruption

Percentage

*Corruption is measured by Transparency International's Corruption Perception Index. "Less Corrupt" countries are those with CPI values in the top third of the range; "Medium Corrupt" in the middle third; and "Highly Corrupt" in the bottom third. We exclude India and China from the population distribution (Figure 1B), but include them in the GDP distribution (Figure 1A).

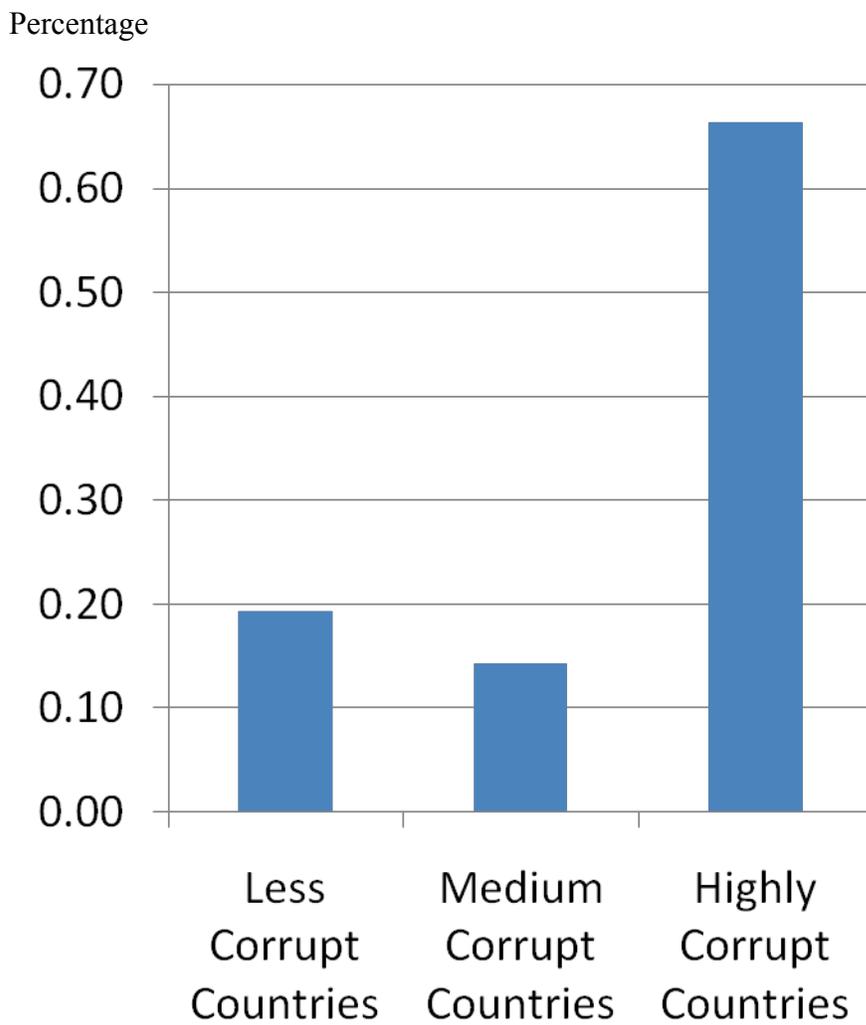Figure 1B. Distribution of World Population by Level of Corruption

Percentage

Table 1
Results of Arizona Classroom Experiment

Treatment (Reported Propensity Untruthful Senders)	Number of Subjects	Percent Truthful	z-statistic (Control - Treatment)	Percent Predicting Receiver Acceptance
Control	97	58.8%		74.2%
Y=15%	25	64.0%	-.480	76.0%
Y=40%	26	53.8%	0.455	80.8%
Y=60%	33	54.5%	0.430	63.6%
Y=85%	52	19.2%	5.349***	73.1%
Overall	233	49.3%		73.4%

*** denotes significant at 1% level (two-sided).

Table 2
Probit Regression of Arizona Sender Message Choices (Truthful vs. Untruthful)
with Course Fixed Effects

Variable	Coefficient	z-statistic	Marginal Effect	z-statistic
Constant	0.7004**	2.06		
Y=85% Treatment	-1.0385***	-3.92	-0.3703***	-4.58
Y=60% Treatment	-0.4358	-1.17	-0.1480	-1.25
Y=40% Treatment	-0.5554	-1.44	-0.1873	-1.58
Y=15% Treatment	0.0206	0.06	0.0073	0.06

Note: N=233. ** denotes significant at 5% (two-sided). *** denotes significant at 1%.
Dependent variable: Sender message choice (truthful=1, untruthful=0).
We report average marginal effects. The fixed course effects are jointly insignificant,
with χ^2 (df=5) test statistic (p-value) 7.43 (.1904); however, one course effect is individually significant.

Table 3
Arizona Survey Results

Treatment	Number Of Subjects	Percent Truthful	z-statistic (Control - Treatment)	z-statistic (Truthful - Untruthful
Control	43	69.8%		
Truthful	63	79.4%	-1.108	
Untruthful	68	50.0%	2.134**	3.707***

, * denotes significant at 5% (**) and 1% (***) levels.

Table 4
Results of Calcutta Lab Experiment

Treatment (Reported Propensity Untruthful Senders)	Number of Subjects	Percent Truthful	z-statistic (Control - Treatment)	Percent Predicting Receiver Acceptance
Control	29	44.8%		82.7%
Y=15%	31	67.7%	-1.836*	74.2%
Overall	60	56.6%		78.3%

* denotes significant at 10% level (two-sided).

Table 5

<u>Probit Regression of Sender Message Choices on Treatments
and Sender Beliefs about Receiver Beliefs</u>

Variable	Coefficient	z-statistic	Marginal Effect	z-statistic
<u>Arizona Subjects</u>				
Constant	-0.0215	-0.04		
Y=85% Treatment	-0.9275**	-2.43	-0.3396**	-2.55
Sender Belief	0.0034	0.45	0.0012	0.45
<u>Calcutta Subjects</u>				
Constant	-0.0980	-0.15		
Y=15% Treatment	0.5927*	1.77	0.2299*	1.82
Sender Belief	-0.0005	-0.05	-0.0002	-0.05

Note: N (for Arizona) = 57 (common course), N (for Calcutta) = 60. Sender beliefs are each Sender's prediction of the Receiver prediction of the proportion of truthful Senders. Dependent variable: Sender message choice (truthful = 1, untruthful = 0).

Supplement to Section 3 of "Is Dishonesty Contagious?" (not for publication)

　　　Inequality Aversion and Contagion. Fehr and Schmidt (1999) assume that (i) some subjects obtain disutility when they are worse off than other subjects and also obtain disutility when they are better off, but (2) suffer more disutility when inequality is to their material disadvantage than when it is to their material advantage. Formally, consider a generalized Fehr-Schmidt subject i utility function,

(A1) $U_i^*(X) = U_i(X_i) - (1/(n-1))\{ \sum_{j=1}^{i-1} \underline{V}_i(X_j - X_i) + \sum_{j=i+1}^{n} \bar{V}_i(X_i - X_j) \} + w_i(\sum_{j=1}^{n} X_j) - \delta a_i,$

where $X=(X_1,\ldots,X_n)$ are the monetary payoffs for the reference group of $j\varepsilon\{1,\ldots,n\}$, whose payoffs (without loss) are ordered from largest to smallest, $X_1 \geq X_2 \ldots \geq X_n$; the direct utility function U is increasing and weakly concave; and the inequity disutility functions $\underline{V}$ and $\bar{V}$ are increasing and weakly convex, with $\underline{V}(0) = \bar{V}(0) = 0$ and $\underline{V}(z) \geq \bar{V}(z)$ for $z>0$ (so that a material disadvantage can yield greater disutility than a material advantage). Weak convexity of the inequity disutility function implies that the marginal disutility of inequity does not decline with the extent of inequity; that is, large inequities are not better, per unit, than small inequities. We add two components to the Fehr-Schmidt utility function, one reflecting a concern for social welfare (with w_i a non-negative constant) and the other an aversion to lying, where $\delta=0$ if agent i is honest, $\delta=1$ if agent i is dishonest, and the parameter a_i is positive. It is easily verified that, with these two additions (and the Sender caring only about his own Receiver), the Fehr-Schmidt preferences of equation (1) are consistent with Gneezy's (2005) findings.

　　　We will assume that (1) the relevant reference group is the set of all participants in the experiment, with n a large number, and (2) a Sender subject i chooses whether to be truthful or untruthful in order to maximize his/her expected utility. Note that the reference group must include a wider population of subjects in order for others' propensity for honesty to have any impact on a Sender's choices.

　　　We denote subject i's perceived probability that other Senders are truthful by p_S, and the probability that Receivers accept their Sender recommendations by p_R. Potential Sender payoffs for the deception game are (X_H^S, X_L^S) and potential Receiver

payoffs are (X_H^R, X_L^R), where H and L denote high and low. For concreteness, we will focus on our Arizona game where $X_H^S = X_H^R = X_H (=6) > X_L^S (=4) > X_L^R (=3)$. Finally, let B_u denote Sender i's net benefit of dishonesty (vs. honesty).

Our experimental results imply the presence of some contagion in dishonesty. Can the Fehr-Schmidt utility function, as interpreted here, explain this contagion? Specifically, does B_u *fall* when the perceived probability of honesty in the reference population (p_S) rises? Differentiating gives (see below for details):

(A2) $\partial B_u / \partial p_S = (2p_R-1)^2 (1/2)[\bar{V}(X_H - X_L^R) - \bar{V}(X_H - X_L^S) - \bar{V}(X_L^S - X_L^R)] \geq 0,$

where the inequality follows from weak convexity of $\bar{V}(.)$ and $X_H > X_L^S > X_L^R$. Hence, the Fehr-Schmidt framework does not predict the contagion in dishonesty that we observe in our Arizona experiment; in fact, it predicts the reverse!

To derive equation (A2) (dropping i subscripts for convenience), let:

p^* = perceived probability that a Sender receives X_H in the reference group

$\quad = (1-p_S)p_R + p_S(1-p_R)$

q^* = perceived probability that a Receiver receives X_H in the reference group

$\quad = p_S p_R + (1-p_S)(1-p_R)$

G_{Li} = expected utility of Sender subject i when i receives X_L^S

$\quad = U(X_L^S) - [(p^*+q^*)/2] \underline{V}(X_H - X_L^S) - [(1-q^*)/2] \bar{V}(X_L^S - X_L^R)$

G_{Hi} = expected utility of Sender subject i when i receives X_H

$\quad = U(X_H) - [(1-p^*)/2] \bar{V}(X_H - X_L^S) - [(1-q^*)/2] \bar{V}(X_H - X_L^R)$

Hence, the net expected utility benefit of untruthful (vs. truthful) conduct is, for large n,

(A3) $B_u = (2p_R-1)(G_{Hi} - G_{Li}) = (2p_R-1) \{[U(X_H) - U(X_L^S)]$

$\quad\quad - (1/2)[\bar{V}(X_H - X_L^S) - \underline{V}(X_H - X_L^S)] - [(1-q^*)/2] \Delta \}$

where we have substituted for $(p^*+q^*) = (1-p^*)+(1-q^*) = 1$;

(A4) $\Delta = \bar{V}(X_H - X_L^R) - \bar{V}(X_H - X_L^S) - \bar{V}(X_L^S - X_L^R) \geq 0,$

and the last inequality follows from weak convexity of $\bar{V}(.)$ and $X_H > X_L^S > X_L^R$. Differentiating gives equation (A2).

In equation (A2), we assume that when a Sender's assessment of p_S changes (with our treatments), his beliefs about p_R – the probability that Receivers accept their Sender recommendations – do not change. As noted above, this premise is consistent with our control for Sender expectations about Receiver behavior and evidence of our success with this control (Tables 1 and 4). Our conclusion here also rests on two key premises: (1) $\bar{V}$ is weakly convex; and (2) the reference group is the population of all subjects (Senders and Receivers), rather than a subset (Senders only). Reversing either of these two assumptions reverses the direction of inequality in equation (A2).

Formally, if $\bar{V}$ is strictly concave, then $\Delta < 0$ in equation (A2) and hence (so long as $p_R \neq 1/2$), $\partial B_u / \partial p_S < 0$ in equation (2). Similarly, if the relevant reference group is *Senders only*, then (assuming $p_R \neq 1/2$),

$$\partial B_u / \partial p_S = (2p_R - 1)(1 - 2p_R)[\underline{V}(X_H - X_L^S) + \bar{V}(X_H - X_L^S)] < 0.$$

However, our premises are natural here. Even under these (plausible) premises, our observations do not imply that Fehr-Schmidt preferences cannot explain contagion in *any* deception game. In our India experiment, for example, we have

$$\partial B_u / \partial p_S \overset{s}{=} \bar{V}(100) - \bar{V}(60) - 2\bar{V}(40) - \underline{V}(40),$$

which is likely to be negative, consistent with the contagion hypothesis.

Guilt Aversion, Sender Beliefs, and Incentives for Dishonesty. Here we establish that guilt aversion implies no clear effect of Sender beliefs on a Sender's decision to be truthful or not. Let $p_S \in [0,1]$ denote the Receiver belief about the probability that the Sender is truthful, $g(p_S; \alpha)$ the probability the Sender assigns to p_S, $p_R(p_S)$ the probability the Sender assigns to Receiver acceptance given p_S (assumed non-decreasing), and U_i the following guilt averse Sender utility function,

$$U_i = X_{Si} + \gamma \min(0, X_{Ri} - E(\tilde{X}_{Ri})),$$

where X_{Si} and X_{Ri} are Sender and Receiver payoffs, E is the Sender's belief about the Receiver's expected payoff, and $\gamma > 0$. Then the Sender's net expected benefit to an

untruthful (vs. truthful) message, given payoff options (X_H^S, X_L^R) and (X_L^S, X_H^R), $X_H^S > X_L^S$, $X_H^R > X_L^R$, is

$$B_u = (X_H^S - X_L^S) \int_0^1 [2p_R(p_S)-1]g(p_S;\alpha)dp_S$$

$$- \gamma (X_H^R - X_L^R)\{ \int_{p_s^*}^1 p_S[2p_R(p_S)-1]g(p_S;\alpha)dp_S + \int_0^{p_s^*} (1-p_S) [2p_R(p_S)-1]g(p_S;\alpha)dp_S\},$$

where $p_s^* = p_S$: $p_R(p_S) > (<) (1/2)$ for $p_S > (<) p_s^*$. We assume that α shifts the p_S distribution upward in the sense of first order stochastic dominance, $G_\alpha(P_S;\alpha) < 0$ for $0 < p_S < 1$, where G is the cumulative distribution function.

Example 1. Suppose $p_R(p_S) = 1$ (0) when $p_S > (<)(1/2)$, so that the Sender expects a rational (self-interested) Receiver decision. Then $p_s^* = 1/2$ and

$$B_u = (X_H^S - X_L^S) [1-2G(1/2;\alpha))] - \gamma (X_H^R - X_L^R)[1-G(1/2;\alpha)- \int_0^1 G(p_S;\alpha)dp_S].$$

Differentiating:

$$\partial B_u/\partial \alpha = - (X_H^S - X_L^S) 2G_\alpha(1/2;\alpha) + \gamma (X_H^R - X_L^R)[G_\alpha(1/2;\alpha) + \int_0^1 G_\alpha(p_S;\alpha)dp_S].$$

The first term is positive and the second is negative. Thus, B_u can rise or fall with the Sender's assessment of Receiver beliefs about Sender propensities for truthfulness.

Example 2. Suppose $p_R(p_S) = p_R > ½$ (constant). Then

$$B_u = (X_H^S - X_L^S) [2p_R-1] - \gamma (X_H^R - X_L^R) [2p_R-1] E(p_S;\alpha),$$

where $E(p_S;\alpha) = \int_0^1 p_S g(p_S;\alpha)dp_S$. Differentiating:

$$\partial B_u/\partial \alpha = - \gamma (X_H^R - X_L^R) [2p_R-1] [\partial E(p_S;\alpha)/\partial \alpha] < 0.$$

Hence, if p_R is constant (as presumed in the discussion in the text), then incentives for truthful behavior rise with Sender beliefs about Receiver expectations of Sender truthfulness.