

Is Dishonesty Contagious?
By Robert Innes and Arnab Mitra
University of Arizona

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 illustrated 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 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. One 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; conversely, high incomes deter corruption which in turn promotes growth.¹

In this paper, we explore another possible contributing explanation that is rooted in individual preferences. Specifically, we conjecture that honesty is *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 attempt to 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 experiment mimics the original deception game designed by Uri Gneezy (2005), who studied the effects of different payoffs on individuals' aversion to untruthful behavior. Unlike Gneezy (2005), we consider a single set of payoffs and focus on the possibility of contagion. In doing this, we find evidence for contagion in the limited

¹ There is a vast literature on the evolution of institutions and their relationship to corruption and growth (see, for example, Acemoglu, et al, 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.

sense that they perception of a particularly strong group propensity for dishonesty promotes untruthful behavior by our subjects to a statistically (and economically) significant extent. We should stress that this result identifies *intrinsic* contagion that is *situation specific*.

The contagion is situation specific in the sense that we are testing for whether individuals alter their behavior in a given situation in response to changes in situation-specific perceptions. For example, one might expect that individuals who come from a culture or peer group that is more dishonest will also tend to be more dishonest, conforming to the social norm for this group. However, if present, this tendency has no bearing on our experiment, because the implied preferences for honesty would *not change* with our treatments.

The contagion is intrinsic in the sense that an individual's decision (to tell the truth or not) is completely private and the stimulated perception of other subjects' behavior (the prevalence of dishonesty) is completely irrelevant to payoffs from the individual's truth-telling decision. As a result, existing explanations for "conformity" and obeying social norms do not apply here. Arguably the most prominent explanation is that there are social or explicit sanctions for violating social norms. Social sanctions can take the form of disapproval or resentment (Sugden, 1998) or reciprocal punishments (see Fehr and Fischbacher, 2004 for an excellent survey). A number of studies document the ability of such sanctions to promote pro-social behavior in, for example, a public goods game (e.g., Fehr and Gächter, 2000, 2002). In our experiment, however, subjects' actions are completely private and there is no sanction whatsoever for any departure from any social norm, whatever it may be.

Other strands of literature explain conformity with others' behavior (that defines a norm) as a mechanism to obtain social esteem (Akerlof, 1980; Bernheim, 1994) or as a way to benefit from information that others have (Banerjee, 1992, Bikhchandaria, Hirshleifer and Welch, 1992). However, acquiring social esteem / status requires that others can observe one's actions, and herd behavior requires that there be relevant information – information about possible payoffs – that can be conveyed by the actions of others. Again, neither is true in our experimental setting: Individual actions are unobservable to anyone other than the individual; and what others do has no bearing on the payoff consequences of individual decisions.

To explain the intrinsic contagion that we identify, one must think about the nature and genesis of social preferences, the topic of a large and growing literature. In general, the key difference between this literature and our paper is that the former identifies and seeks to explain why and how an individual may care about other players *who are affected by* one's actions, as opposed to how this social preference is affected by seemingly irrelevant information about other players' decisions. For this and other reasons, we argue in Section 3 below that a variety of ingenious explanations for experimental results in fairness / dictator, ultimatum, market, trust and other games (see, for example, Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002; Charness and Dufwenberg, 2006) are unlikely to explain the intrinsic contagion that we observe. Our conjecture, therefore, is that contagion may be a hard-wired attribute of individual preferences. This leads us to inquire, in a companion paper, about potential evolutionary motives for in-bred

contagion, drawing broadly on the evolutionary strategy literature (e.g., Frank, 1987; Bergstrom, 2002; Guth and Kliemt, 1994; Sobel, 2005).²

Ours is not the first paper to study contagion per se, although it is the first (to our knowledge) to study contagion in dishonesty. Most closely related is a key (and independent) paper by Bicchieri and Xiao (BX, 2007), who study a dictator game wherein 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.³ However, a key distinction between this result and ours (in addition to the important difference in context, dictator vs. deception) concerns the role of dictator / sender beliefs about receiver expectations. While we control for these beliefs – and have evidence of success in doing so – BX elicit these beliefs and find that their treatments affect them. That is, when dictators are told that prior dictators tended to be more fair, they also believe that receivers expect dictators to be more fair.⁴ Therefore, interestingly, BX’s contagion results can be explained by subject guilt aversion, as posited by Charness and Dufwenberg (2006). Our results cannot, as we discuss in Section 3 below.

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. Finally, Fisman and Miguel (2007) famously study the tendency for diplomats to garner parking tickets in New York, finding that the immunity-protected foreigners take their home country propensities for lawlessness with them. While these results could conceivably be interpreted as evidence against contagion (because diplomats seem to ignore U.S. values in their behavior), we believe that such inferences are misplaced for two reasons. First, there is no *ceteris paribus* in this comparison; diplomats may well temper their lawless

² In the companion paper, we build on literature which explains “conformity” as a symptom of network externalities (for example, Katz and Shapiro, 1986; Banerjee and Besley, 1990). Although our experiment contains no network effects, we believe that these arguments nonetheless offer the best hope for explaining the contagion that we observe. If many situations that involve decisions on truth and/or trust and/or fairness embed network effects, then having a rule or routine embedded in preferences – telling an individual to “do as others do” in these types of situations – may be advantageous from an evolutionary perspective. While we explore this logic in the companion paper, we note that the existing literature does not motivate contagious preferences that *vary* with peer group propensities (for honesty, in our context). Rather they motivate conformist equilibria in which individuals do as others do *in each equilibrium*.

³ The results of BX are somewhat mixed in the following respect. On one hand, the proportion of dictators who are fair vs. selfish when told prior dictators were predominantly fair (FC) vs. predominantly selfish (SC) are insignificantly different from one another. However, in a probit regression, they find that there is a significant positive effect of a dictator’s expected proportion of other dictators’ fair (vs. selfish) allocations on the dictator’s own likelihood of choosing fairly.

⁴ BX elicit dictator beliefs about the proportion of receivers who expect a fair (vs. selfish) allocation. Under BX’s FC (fair) treatment, the average reported dictator belief about receiver expectations of fairness is approximately 48 percent (with approximate standard deviation of .06 and 23 observations); and under the SC (selfish) treatment, it is approximately 17 percent (with approximate standard deviation of .045 and 21 observations). The difference between the two averages is clearly large and statistically significant.

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.

2. Experimental and Survey Evidence

In this paper, we are interested in whether and how one's perception of others' propensity to be honest or dishonest affects one's own choices of whether or not to be honest, even when the behavior of others has no material bearing on the outcomes from one's own choices and individual decisions are completely private.

A. The 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.

In our experiment, we focus on a single set of payment options (while 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 \$6.⁵

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:

⁵ 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.

“Out of a sample 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). The statements reported in the treatments are literally truthful; however, the reported samples are of course not random.⁶

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 also 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.”

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

To verify that Senders generally believed that Receivers would accept their recommendations, we followed Gneezy’s (2005) approach, asking them to predict their Receiver’s choice and paying them for a correct prediction. Overall, 76.7% of Senders predicted that their Receiver would accept their recommendation.⁷ These results are instructive for a couple of reasons. Because Senders generally expect their recommendations to be followed, their choices reflect a concern for the “fairness” / morality of lying, and not strategic motives. In addition, this expectation argues against the guilt aversion hypothesis (Charness and Dufwenberg, 2006) as an explanation for our results (more on this subject in a moment). As it turned out, 71 percent of our Receivers followed their Sender recommendations.

The experiment was conducted in undergraduate economics classes at the University of Arizona in Spring, 2008. In total, there were 176 Senders and 176

⁶ A norm in experimental economics is that the experimenter be honest to his/her subjects. We obey this norm with our approach. However, our need in this experiment to influence perceptions requires an element of manipulation. We note that such manipulations are common in the experimental economics literature. Prominent examples are influential papers 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 Riley, 2002), and that use the standard experimental protocol to not inform subjects that they will be playing more than one experiment or experimental round (e.g., Binmore, et. al, 1985). See Bonetti (1998) for a lucid discussion of this topic.

⁷ 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 accept decisions (77.4%) is essentially identical to the proportion of untruthful Senders predicting accept decisions (75.9%) in our experiment.

Receivers. Receivers were in a different class 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.⁸

Table 1 reports the number of Senders exposed to each of the different treatments, and summarizes our results. Relative to the control, one treatment has a significant effect (in both statistical and economic senses) 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 83 percent (under the treatment), more than doubling. Other treatments have no significant effect. We thus find support for the contagion hypothesis in the limited sense that a strong propensity for untruthfulness is contagious.

B. Survey Evidence

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:

"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 2 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 at least the limited sense that a perceived propensity for others to be highly untruthful is contagious.

⁸ 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.

3. Explanations for Contagion: Discussion

Our results indicate that some of our subjects have an aversion to lying that depends positively on the perceived propensity for honesty in a relevant peer group. Why might this be true?

A number of papers posit inter-dependent preferences that can explain other otherwise puzzling experimental findings. Bolton and Ockenfels (2000) posit preferences that depend upon one's relative payoff in a peer group. Charness and Rabin (2002) consider a variety of possible forms for inter-dependent preferences, including a component that weights social welfare and reciprocal preferences that give positive utility to punishment of "bad behavior" by partners and to rewards of "good behavior." In the context of our experiment, individual preferences that weight social welfare cannot explain contagion; regardless of what other agents do, the impact of a Sender's choice on social welfare (the overall payoff from the experiment) is the same. Similarly, there is no real role for reciprocity in the sense that we control for beliefs about Receiver behavior.

In this Section, we consider two theories that offer the most promise as possible explanations for our results. First is the "guilt aversion" hypothesis of Charness and Dufwenberg (2006), which recognizes the potential importance of Sender beliefs about Receiver expectations (drawing upon the early work of Rabin, 1993). Second, a concern for relative payoffs (Bolton and Ockenfels, 2000) and inequity (Fehr and Schmidt, 1999) could lead to effects of perceived group behavior on individual behavior. We focus on Fehr and Schmidt's (1999) inequality aversion (which is also studied by Charness and Rabin, 2002); similar logic applies to the ERC preferences of Bolton and Ockenfels (2000).

We argue that neither of these ingenious theories offers a convincing explanation for our results. This conclusion suggests that the "moral contagion" that we find is likely to be a hard-wired attribute of individual preferences.

A. Guilt Aversion

Charness and Dufwenberg (CD, 2006) argue that subjects are averse to disappointing their partners; that is, 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.

In our experiment, this mechanism could be important if (and only if) our treatments affect a Sender's beliefs about the Receiver's expectations. For example, 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.

There are at least three reasons why this logic is unlikely to explain our results. First, this logic presumes that Senders assume that Receivers base expectations of payoffs on the true payoffs available in the game. However, the Receivers in our experiment are never told the payoffs available in the game and,

hence, have no basis for disappointment. Senders are told this in the experiment instructions. Second, the Sender knows that the information about Sender behavior in prior sessions – our treatment – is not provided to Receivers. Third, evidence that Senders do not in fact revise their beliefs about Receiver expectations, in response to our treatments, is provided by their predictions of Receiver decisions. If Senders believe that Receivers somehow internalize the information that Senders are heavily untruthful (our treatment), they should also expect Receivers to revise their decisions on whether or not to accept or reject Sender recommendations. In particular, we would expect to see more predictions that Receivers reject their Sender recommendations under the “heavily untruthful” treatment (vs. our control). However, as documented in Table 1, we find no such effect. Indeed, restricting attention to Senders who predict Receiver acceptances of their recommendations, the impact of our “heavily untruthful” (Y=85%) treatment is even larger than for the full sample.

Turning to our survey, we again believe that CD’s “guilt aversion” is unlikely to explain our results. First the counter argument. A buyer’s expected loss, when he/she is not informed (vs. informed) about the car problem, rises with the societal propensity to be honest. The reason is that, with a higher societal propensity for honesty (call it h), the buyer expects the seller to report any problem with higher probability, leading to a lower expected probability that the car has a problem when no information to this effect is forthcoming from the seller; hence, the buyer’s discovery of a problem is more of a surprise when h is higher. All of this suggests that the Seller expects greater buyer disappointment from deception when h is higher, which would seem to explain our survey results.⁹ However, in our hypothetical example, we *fix* the expected loss to the buyer, from seller dishonesty, at \$250 *in all treatments*. Hence, even though the societal propensity for honesty h affects “guilt” penalties in an equilibrium, our treatments do not affect this penalty in the posed situation.

⁹ For example, suppose that the car has at most one problem with probability p , which (if it exists) is known by the seller with probability q . Then, given a perceived propensity for the seller to report the problem (if he/she is aware of it) h , the probability of a problem given that the seller does not report one is:

$$p^* = p(1-q) + pq(1-h) = p(1-qh),$$

namely, the probability that there is a problem of which the seller is not aware, plus the probability that there is a problem known but unreported by the seller. The associated expected loss to the buyer when the seller does not report a (known) problem is:

$$\text{Loss} = c(1-p^*) = c(1-p+pqh),$$

where c is the cost of dealing with the problem. Clearly, this loss rises with h , provided q is positive. However, as we argue in the text, we fix this expected loss at \$250 across all treatments in our hypothetical example. Specifically, consider the price that the buyer is willing to pay for the car under the two alternate seller strategies (report or not report), where V is the buyer’s “base price” without a possible problem. If the seller does not report (NR), then the price is reduced by the expected cost of the problem,

$$P_{NR} = V - p^*c.$$

If the seller reports, then the price is reduced by the full cost of the problem,

$$P_R = V - c.$$

The reduction in price from reporting the problem, which we fix in our example at \$250, is

$$P_R - P_{NR} = -c(1-p^*),$$

which is exactly the expected loss to the buyer when the seller does not report.

B. Inequality Aversion

Fehr and Schmidt (1999) explain a wide variety of experimental results by positing that subjects are averse to unequal outcomes in a reference group. Specifically, they 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 the generalized Fehr-Schmidt subject i utility function,

$$(1) \quad U_i^*(X) = U(X_i) - (1/(n-1)) \left\{ \sum_{j=1}^{i-1} \underline{V}(X_j - X_i) + \sum_{j=i+1}^n \bar{V}(X_i - X_j) \right\}$$

where $X=(X_1, \dots, X_n)$ are the monetary payoffs for the reference group of $j \in \{1, \dots, n\}$, whose payoffs (without loss) are ordered from largest to smallest, $X_1 \geq X_2 \dots \geq X_n$; 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.¹⁰

Let me first consider implications of the Fehr-Schmidt utility function for the deception game studied in this paper. Let us further suppose 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.¹¹

We denote subject i 's perceived probability that other Senders are truthful by p_S and that Receivers accept their Sender recommendations by p_R , and we assume that $p_R > 1/2$. 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 (from the game studied above) $X_H^S = X_H^R = X_H (=6) > X_L^S (=4) > X_L^R (=3)$. Now note:

$$\begin{aligned} p^* &= \text{perceived probability that a Sender receives } X_H \text{ in the reference group} \\ &= (1-p_S)p_R + p_S(1-p_R) \\ q^* &= \text{perceived probability that a Receiver receives } X_H \text{ in the reference group} \\ &= p_S p_R + (1-p_S)(1-p_R) \\ G_{Li} &= \text{expected utility of Sender subject } i \text{ when } i \text{ receives } X_L^S \\ &= U(X_L^S) - [(p^*+q^*)/2] \underline{V}(X_H - X_L^S) - [(1-q^*)/2] \bar{V}(X_L^S - X_L^R) \\ G_{Hi} &= \text{expected utility of Sender subject } i \text{ when } i \text{ receives } X_H \end{aligned}$$

¹⁰ Bolton and Ockenfels (2000) make a similar assumption, their Assumption 3.

¹¹ In order for the perceived propensities for honesty in the population of experimental participants to have any impact on a Sender's behavior, the reference group must include those in this wider population, as assumed here. We should note, however, that in the applications that they consider, Fehr and Schmidt (1999) do not envision this broad reference group, instead restricting attention to those that are directly affected by an agent's decision (proposer and responder). Our premise on this score is controversial. For example, Ellingen and Johannesson (2008, p. 994) express skepticism that agents care about the outcomes from others' actions in choosing their own conduct.

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

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

$$(2) \quad B_u = (2p_R - 1)(G_{Hi} - G_{Li}) = (2p_R - 1) \{ [U(X_H) - U(X_L^S)] - (1/2)[\bar{V}(X_H - X_L^S) - \underline{V}(X_H - X_L^S)] - [(1-q^*)/2] \Delta \}$$

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

$$(3) \quad \Delta = \bar{V}(X_H - X_L^R) - \bar{V}(X_H - X_L^S) - \bar{V}(X_L^S - X_L^R) \geq 0,$$

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

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 Sender subject i 's net benefit of dishonesty, B_u , *fall* when the perceived probability of honesty in the reference population (p_S) rises?

Differentiating (2),

$$(4) \quad \partial B_u / \partial p_S = (2p_R - 1)(1/2)(\partial q^* / \partial p_S) \Delta \geq 0,$$

where the inequality follows from $p_R > 1/2$, $\Delta \geq 0$ (equation (3)) and

$$\partial q^* / \partial p_S = (2p_R - 1) > 0.$$

Hence, the Fehr-Schmidt framework posited here cannot explain contagion in dishonesty.

Notably, this conclusion rests on two key premises made here:¹² (1) \bar{V} is weakly convex; and (2) the relevant reference is the population of all participants in the experiment (Senders and Receivers), rather than a subset (e.g., Senders only). Reversing either of these two assumptions leads to a reversal in direction of

inequality in equation (4). If \bar{V} is strictly concave, then $\Delta < 0$ in equation (3) and hence (so long as $p_R \neq 1/2$), $\partial B_u / \partial p_S < 0$ in equation (4). 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, we believe our premises are plausible. First consider the reference group. Virtually all scholars investigating inter-dependent preferences in experiments assume that a proposer (Sender) cares about his responder (Receiver). If one is to believe that the realm of inter-dependency of preferences is broader, encompassing other individuals in the experiment, then why would one's own responder matter and no others? Second consider the weak convexity premise. Concavity would imply that an individual is less averse to larger, vs. smaller, inequalities. For example, consider two payoff distributions among two players, Self and Other, (6,1) and (6,3). Under inequality aversion, the first is clearly worse than the second. Now consider Self's willingness to pay for a given reduction in inequality, namely, the amount y such that Self is indifferent between (1) keeping the initial allocation ((6, X_0), with $X_0 = 1$ or 3), and (2) giving up y in exchange for a one dollar increase in Other's payoff:

¹² If $p_R \leq 1/2$, it is easily seen that the inequality in (4) persists. Hence, the above conclusion does not depend upon the (plausible) premise that $p_R > 1/2$.

$$U_{self}^* (X_{Self}=6-y, X_{Other}=X_0+1) = U_{self}^* (X_{Self}=6, X_{Other}=X_0)$$

Our convexity assumption implies that Self is willing to give up more (so that y is larger) when the inequality is worse (when X_0 is 1 rather than 3). Concavity implausibly implies the opposite.

4. Conclusion

We study whether subjects propensity to lie in a deception game depends upon the perceived propensity for others to lie. We find evidence of such “contagion” in the limited sense that when subjects are led to believe that a large majority of their peers are untruthful, they are more likely to be untruthful themselves. We argue that these results are unlikely to be a symptom of the other-regarding preferences proposed in current theory (e.g., Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002; Charness and Dufwenberg, 2006). To explain our results, we therefore believe that one should inquire about whether moral contagion, as an evolutionary trait, might be advantageous. That is, perhaps individuals have hard-wired rules or routines that tell them to revise their preferences in response to social cues of the type that we provided in our experiment. Their hard-wired response to perceptions that dishonesty is predominant is then to drop their aversion to dishonesty. The question becomes: Why does such a hard-wired response create an evolutionary advantage? That is, why should this “contagion” trait evolve? In a companion paper (Innes, 2008), I argue that such contagion may be advantageous in overcoming barriers to cooperation. When most people are dishonest, honesty creates a barrier to cooperation because honest people are exploited and therefore withdraw from joint ventures. Conversely, when most people are honest, dishonesty creates a barrier to cooperation because honest people refuse to work with dishonest people. In both cases, “copying the majority” is advantageous. We conjecture that our results are likely an outgrowth of such logic.

References

- Acemoglu, D., S. Johnson, and J. Robinson. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review* 91 (2001): 1369-1401.
- Akerlof, G. "A Theory of Social Custom, of Which Unemployment May Be One Consequence." *Quarterly Journal of Economics* 94 (1980): 749-75.
- Banerjee, A. "A Simple Model of Herd Behavior." *Quarterly Journal of Economics* 107 (1992): 797-817.
- Banerjee, A., and T. Besley. "Peer Group Externalities and Learning Incentives: A Theory of Nerd Behavior." John M. Olin Discussion Paper no. 68. Princeton, N.J.: Princeton University, December 1990.
- Bergstrom, T. "Evolution of Social Behavior: Individual and Group Selection." *Journal of Economic Perspectives* 16 (2002): 67-88.
- Bernheim, B.D.. "A Theory of Conformity." *Journal of Political Economy* 102 (1994): 841-77.
- Bertrand, M., and S. Mullainathan. "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94 (2004): 991-1013.
- Bicchieri, C., and E. Xiao. "Do the Right Thing: But Only if Others Do So." University of Pennsylvania Working Paper, Philosophy, Politics and Economics Program, August 2007.
- Bikhchandaria, S., D. Hershleifer, and I. Welch. "A Theory of Fads, Fashion, Custom, and Cultural change as Informational Cascades." *Journal of Political Economy* 100 (1992): 992-1026.
- Binmore, K., A. Shaked, and J. Sutton. "Testing Noncooperative Bargaining Theory: A Preliminary Study." *American Economic Review* 75 (1985): 1178-80.
- Bolton, G. and A. Ockenfels. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review* 90 (2000): 166-93.
- Bonetti, S. "Experimental Economics and Deception." *Journal of Economic Psychology* 19 (1998): 377-95.
- Charness, G., and M. Dufwenberg. "Promises and Partnership." *Econometrica* 74 (2006): 1579-1601.
- Charness, G. and M. Rabin. "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics* 117 (2002): 817-69. 2002.
- Ellingsen, T., and M. Johannesson. "Pride and Prejudice: The Human Side of Incentive Theory." *American Economic Review* 98 (2008): 990-1008.
- Fehr, E., and U. Fischbacher. "Social Norms and Human Cooperation." *Trends in Cognitive Sciences* 8 (2004): 185-90.
- Fehr, E., and S. Gächter. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review* 90 (2000): :980-94.
- Fehr, E., and S. Gächter. "Altruistic Punishment in Humans." *Nature* 415 (2002): 137-40.
- Fehr, E. and K. Schmidt. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics* 114 (1999): 817-68.

- Fisman, R., and E. Miguel. "Corruption, Norms and Legal Enforcement: Evidence From Diplomatic Parking Tickets." *Journal of Political Economy*, 115 (2007): 1020-48.
- Frank, R. "If Homo Economicus Could Choose His Own Utility Function, Would He Want One with a Conscience?" *American Economic Review* 77 (1987): 593-604.
- Gneezy, U. "Deception: The Role of Consequences." *American Economic Review* 95 (2005): 384-94.
- Guiso, L., P. Sapienza, and L. Zingales. "The Role of Social Capital in Financial Development." *American Economic Review* 94 (2004): 526-56.
- Guth, W. and H. Kliemt. "Competition or Cooperation: On the Evolutionary Economics of Trust, Exploitation and Moral Attitudes." *Metroeconomica* 45 (1994): 155-87
- Katz, M., and C. Shapiro. "Technology Adoption in the Presence of Network Externalities." *Journal of Political Economy* 94 (1986): 822-41.
- Lindbeck, A., S. Nyberg, and J. Weibull. "Social Norms and Economic Incentives in the Welfare State." *Quarterly Journal of Economics* 114 (1999): 1-35.
- List, J., and D. Reiley. "The Effects of Seed Money and Refunds on Charitable Giving: Experimental Evidence from a University Capital Campaign." *Journal of Political Economy*, 110 (2002): 215-33.
- Rabin, M. "Incorporating Fairness into Game Theory and Economics." *American Economic Review* 83 (1993): 1281-1302.
- Sobel, J. "Interdependent Preferences and Reciprocity." *Journal of Economic Literature* 43 (2005): 393-436.
- Sugden, R.. "Normative Expectations: The Simultaneous Evolution of Institutions and Norms." In: Ben-Ner, A. and Putterman, L. (eds.), *Economics, Value, and Organization*. Cambridge: Cambridge University Press, 1998.
- Zak, P. and S. Knack. "Trust and Growth." *Economic Journal* 111 (2001): 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 a sample 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): ****

Message A

Message B

(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

Option A

I believe my Receiver will choose

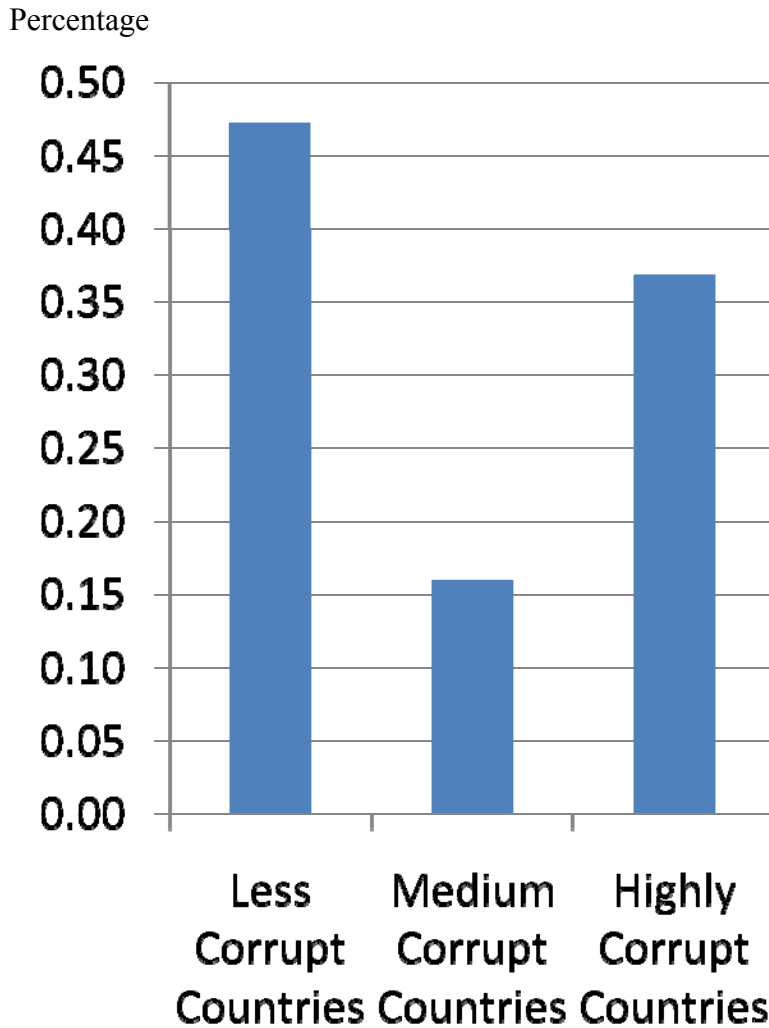
Option B

Your gender (please circle one):

Male

Female

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



*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 Corrupton

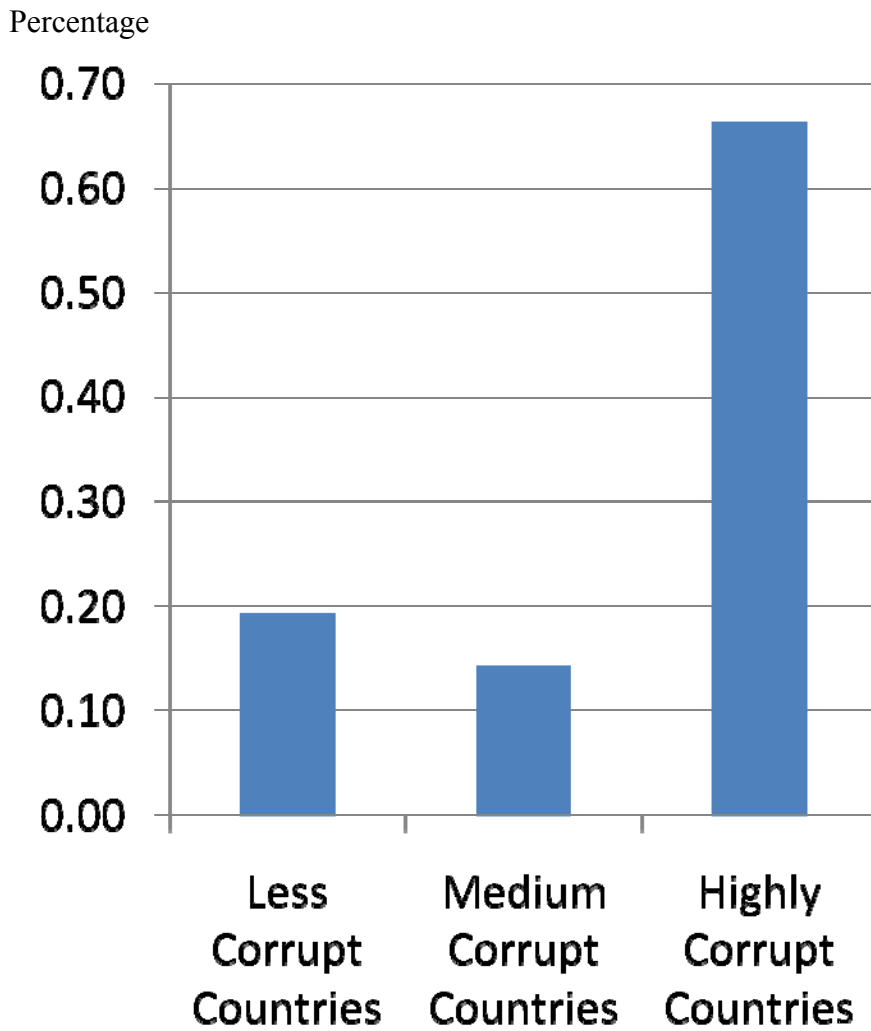


Table 1
Results of Classroom Experiment

Treatment (Reported Propensity Untruthful Senders)	Number of Subjects	Percent Truthful	z-statistic (Control - Treatment)	Percent Predicting Receiver Acceptance	Percent Truthful for Accept Predictors	z-statistic (Control - Treatment)
Control	69	59.4%		76.8%	60.4%	
Y=15%	25	64.0%	-0.406	73.1%	63.2%	
Y=40%	26	53.8%	0.488	80.8%	57.1%	
Y=60%	33	54.5%	0.465	63.6%	61.9%	
Y=85%	23	17.4%	4.258***	91.3%	14.3%	4.532***

*** denotes significant at 1% level.

Table 2
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.