

Journal Club

Editor's Note: These short, critical reviews of recent papers in the *Journal*, written exclusively by graduate students or postdoctoral fellows, are intended to summarize the important findings of the paper and provide additional insight and commentary. For more information on the format and purpose of the Journal Club, please see http://www.jneurosci.org/misc/ifa_features.shtml.

Suboptimal Dishonesty: Rationality in the Absence of Strategic Behavior in Honesty Experiments

 Michele Fioretti and Sean Marden

Department of Economics, University of Southern California, Los Angeles, California 90089
Review of Abe and Greene

What compels honesty or dishonesty in individual decision-making? The “moral grace” hypothesis offers a possible explanation, according to which people are innately honest. In contrast, the “will” hypothesis suggests that people actively suppress dishonest behaviors, which arise naturally. A study by Abe and Greene (2014) has reconciled these two seemingly inconsistent theories by showing that the drivers of honesty depend on the desirability of the reward.

Under the pretense of an experiment on clairvoyance, participants in Abe and Greene (2014) were asked to predict future outcomes of randomized coin flips in an fMRI scanner. Their rewards depended on the accuracy of those predictions and accuracy was self-reported, thus providing both incentive and opportunity for lying. Subjects were classified based on a conservative measure of deviations from chance as “honest” (mean accuracy = 50.1%), “dishonest” (83.6%), or “ambiguous” (67.1%). The authors measured reward desirability by activation of the nucleus accumbens to anticipated rewards, and found a positive correlation between activity in this region and dishonest behavior. They also tested whether

higher activity in the dorsolateral prefrontal cortex (DLPFC) is associated with dishonest subjects behaving honestly, as a proxy for will. Consistent to this, they found significant responses in the DLPFC only for subjects in the ambiguous and dishonest groups when refraining from dishonesty. These combination of results led Abe and Greene (2014) to conclude that honesty flows automatically in presence of weak responses to anticipated rewards, while “will” is needed to refrain from dishonest behaviors in cases of high relative responses.

The choice between responding truthfully and lying that participants faced in the experiment is a strategic decision that can be examined in a game theoretical context. Evaluating their results under this framework may suggest interpretations of participants' behavior or extensions for further studies, which may not have otherwise been considered. *Homo economicus* would expect the subjects to report all of their predictions as accurate; however, in this experiment, not even the “dishonest” group reported perfect accuracy, despite the fact that doing so would guarantee the highest payoffs.

That participants fail to optimize their payoffs does not undermine rationality, it merely indicates that their utility depends on variables beyond monetary returns. Since, in each trial, the potential reward is a fixed amount, it is natural to focus on potential costs associated with lying. A growing body of empirical work points not only to there being a cost incurred

from lying, but also that the subjective cost is heterogeneous across individuals (Gneezy, 2005). There are some participants who choose to tell the truth even when doing so comes at a high opportunity cost, and others who choose to lie when there is no benefit to doing so (Gibson et al., 2013). Typically, these studies have focused on shame aversion, guilt aversion, and risk aversion as potential sources of this cost to lying, but finding meaningful measurements for these emotions can be complex, especially when it comes to disentangling them. Importantly, even taking an agnostic view on the sources of the cost can still elucidate the mechanisms underlying (dis)honesty. Integrating the empirical findings of heterogeneous cost as an *a priori* assumption, subjects showing a spectrum of accuracies when facing identical financial reward (as seen in Abe and Greene, 2014) becomes an expected result of rational individuals maximizing their utility. Still, there is much to be gained from attempts to refine the understanding of the basis and magnitude of this cost.

Strategic interactions between subjects is useful in capturing relevant situations for lying behaviors and thus useful for assessment. These often rely on second-order beliefs (what the sender believes about the beliefs of the receiver), as in Gneezy (2005), which can complicate measurement. To alleviate this concern, another method considers responses given by a subject to the experimenter (Abe and Greene, 2014). In a similar study by Mazar

Received Oct. 15, 2014; revised Dec. 4, 2014; accepted Dec. 5, 2014.

We thank Prof. Giorgio Coricelli for his guidance and encouragement in our research.

Correspondence should be addressed to Michele Fioretti, Department of Economics-KAP 300, University of Southern California, 3620 S. Vermont Ave., Los Angeles, CA 90089-0253. E-mail: fioretti@usc.edu.

DOI:10.1523/JNEUROSCI.4268-14.2015

Copyright © 2015 the authors 0270-6474/15/351817-02\$15.00/0

et al. (2008), two groups of students were asked to answer a mathematical quiz. The control group was graded by the experimenter, while the treatment group was asked to correct their own answers. Notably, the latter group had 10% more correct answers than the control group, on average. In another study, participants were asked to roll a die and subsequently rewarded on a scale monotonically increasing in outcome, with the warning that they would get nothing by reporting a 6 (Fischbacher and Föllmi-Heusi, 2013). The experimenter could not verify the actual outcome, but could use its distribution to assess dishonest behaviors. Game theory suggests that lying behavior would result in the reporting of a 5, yet the resulting distribution also included over-reporting of 4, indicating suboptimal dishonesty. While the subject-experimenter paradigm potentially reduces the effects of second-order beliefs, there are still strategic interactions between the participants and the experimenter.

There is certainly the potential for such strategizing in Abe and Greene (2014), where participants were informed that the study was about how monetary incentives affect clairvoyance. Students who noticed cheating was possible were told that the ability to do so was a “necessary by-product of the experimental design and were encouraged to follow the directions, which preclude cheating” (Abe and Greene, 2014). Even without this disclaimer, subjects may incur guilt over “ruining” the experiment by lying about their prediction ability. As a result, students could be playing a sort of hide-and-peek game by discarding dishonest opportunities if the return is not large enough. While this was ruled out for the “honest” subjects (through comparison of final winnings to a simulated honest subject), the “ambiguous” and “dishonest” groups were not examined in the same way. This may be an erroneous exercise for the “dishonest” group, as their behavior is of interest for other reasons, but a study of the “ambiguous” group may provide valuable insights.

Economic theory posits that the most sophisticated individuals should behave dishonestly all the time, but under this framework the most sophisticated might behave ambiguously to maximize their utilities by camouflaging as honest. There is recent evidence suggesting that subjects with higher IQ base their decisions in a current trial on a greater number of past trials than subjects with low IQ (Hawes et al., 2014). Those who wish to strategically underreport their accuracy on trials with

low-value outcomes would require the sophistication of basing their decisions on a great number of trials. Since they are more likely to have high IQ (Hawes et al., 2014), and high IQ translates—at least theoretically—to dishonest actions in this game, the “ambiguous” group may exhibit the “strategic underreporting” that was shown to not occur in the “honest” group (Abe and Greene, 2014). This may also contribute to the higher mean response times seen in the “ambiguous” group compared with both the “honest” and “dishonest” groups in both the win and loss trials, as well as the slower reaction times observed during dishonest behavior (Abe and Greene, 2014). Also, in a result from Hawes et al. (2014), subjects with higher IQ exhibited weaker striatal BOLD signals after the reward was received, which is not inconsistent with the results in Abe and Greene (2014). This shows that sophistication has a role in the biology of decision-making (Coricelli and Nagel, 2009). Cognitive hierarchy theory (Camerer et al., 2004) corroborates the idea that IQ positively correlates with strategic behaviors; future extensions in this field could help us better understand the ambiguous group.

Investigating the patterns of lying throughout the duration of the experiment could also shed light into dishonesty, particularly in the ambiguous group. In a study by Gneezy et al. (2013), lying increases over time within trials with high benefits arising from dishonest behaviors (lying more than doubles). This implies an attraction learning mechanism, allowing individuals to maximize their payoffs more often as deviation from honest behaviors becomes more lucrative. Camerer et al. (2002) suggests that individuals believing that others are learning may change their actions accordingly, providing one more reason to behave dishonestly. Another explanation, given by Gino et al. (2011), concludes that self-control is needed for truth-telling, but this is not effortless: fatigued students end up behaving dishonestly not because they have a preference for dishonesty but because restraining from doing so is too expensive. So, people in the “ambiguous” group may be those for whom learning to play strategically takes longer, or those who do not tire as quickly, and hence have mean accuracy lower than the dishonest group and higher than the honest group. This would clearly bias our measure of lies. Therefore, as an extension to Abe and Greene (2014), future studies should analyze the pattern of students’ actions with

regards to physical markers and in relation to their perception of peer performance, in order to have an accurate understanding of the drivers of choice at the individual level.

Abe and Greene’s (2014) evidence supports a reconciliation of the “Grace” and “Will” hypotheses wherein a natural state of “Grace” results in honesty for low neural stimuli, and “Will” when sufficiently high neural responses to anticipated rewards (Abe and Greene, 2014), but it also supports a simple utility maximization problem with a heterogeneous cost to lying. Future research should examine and single out different sources of this cost to dishonesty. Several of the behavioral games discussed earlier could be replicated in an fMRI setting to target neural activity when guilt aversion or shame aversion is isolated. Approaching the problem from this context may help in finding a definitive reason why people eschew their best alternative even when doing so causes no evident harm.

References

- Abe N, Greene JD (2014) Response to anticipated reward in the nucleus accumbens predicts behavior in an independent test of honesty. *J Neurosci* 34:10564–10572. [CrossRef Medline](#)
- Camerer CF, Ho TH, Chong JK (2002) Sophisticated experience-weighted attraction learning and strategic teaching in repeated games. *J Econ Theory* 104:137–188. [CrossRef](#)
- Camerer CF, Ho TH, Chong JK (2004) A cognitive hierarchy model of games. *Q J Econ* 119:861–898. [CrossRef](#)
- Coricelli G, Nagel R (2009) Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proc Natl Acad Sci U S A* 106:9163–9168. [CrossRef Medline](#)
- Fischbacher U, Föllmi-Heusi F (2013) Lies in disguise: an experimental study on cheating. *J Eur Econ Assoc* 11:525–547. [CrossRef](#)
- Gibson R, Tanner C, Wagner AF (2013) Preferences for truthfulness: heterogeneity among and within individuals. *Am Econ Rev* 103:532–548. [CrossRef](#)
- Gino F, Schweitzer ME, Mead N, Ariely D (2011) Unable to resist temptation: how self-control depletion promotes unethical behaviour. *Organ Behav Hum Dec Process* 115:191–203. [CrossRef](#)
- Gneezy U (2005) Deception: the role of consequences. *Am Econ Rev* 95:384–394.
- Gneezy U, Rockenbach B, Serra-Garcia M (2013) Measuring lying aversion. *J Econ Behav Organ* 93:293–300. [CrossRef](#)
- Hawes DR, DeYoung CG, Gray JR, Rustichini A (2014) Intelligence moderates neural responses to monetary reward and punishment. *J Neurophysiol* 111:1823–1832. [CrossRef Medline](#)
- Mazar N, Amir O, Ariely D (2008) The dishonesty of honest people: a theory of self-concept maintenance. *J Market Res* 45:633–644. [CrossRef](#)