The slippery slope of dishonesty

Jan B Engelmann & Ernst Fehr

Recent experiments suggest that dishonesty can escalate from small levels to ever-larger ones along a 'slippery slope'. Activity in bilateral amygdala tracks this gradual adaptation to repeated acts of self-serving dishonesty.

Well, you know what happens is, it starts out with you taking a little bit, maybe a few hundred, a few thousand. You get comfortable with that, and before you know it, it snowballs into something big. —Bernard Madoff, as recounted by his secretary in *Vanity Fair*, 2009

Honesty is a fundamental value that is ranked highly across many societies around the globe[1]. Nevertheless, research in behavioral economics and psychology has repeatedly observed dishonest behavior when individuals can benefit from their dishonesty[2], and this pattern persists in many societies[3]. Usually, the level of dishonesty reported in these experiments is relatively low, with participants cheating only about 20% of the maximal possible level of dishonesty even when there is little chance of being caught[4]. Recently, however, we have witnessed a number of striking and widely covered examples of dishonesty, ranging from Ponzi schemes to publicly misrepresenting earnings (see Enron) to disguising the true risk of financial assets (many banks in the lead-up to 2009), all of which contribute to a dishonest business culture[5] that may have partially caused the 2009 financial crisis. One intriguing hypothesis suggests that such crass cases of dishonesty start with an initially small, easily justifiable transgression that gradually escalates to ever-larger ones[6]. Results from a recent functional magnetic resonance imaging (fMRI) experiment by Garrett *et al.*[7] show that the blood oxygenation level–dependent (BOLD) signal in the amygdala gradually adapts to self-serving dishonesty, suggesting that amygdala adaptation may represent the neurobiological basis of the slippery slope of moral degradation.

The authors invited participants to play a two-party task. One subject, the estimator, who was in fact a confederate, was asked to repeatedly guess the amounts of money inside glass jars filled with one-penny coins, shown in photographs. The other subject was placed inside an MRI scanner and given the role of advisor, with the task of providing recommendations about the amount of coins inside the jar. The advisor's advantage over the estimator was that he or she was presented with a larger and better-resolution image of the jar and also knew the possible range of monetary amounts inside the jar.

The advisor was led to believe that both players of the game were paid based on the estimator's accuracy. However, supposedly unbeknownst to the estimator, the incentive structure for the advisor changed during the experiment to measure the evolution of dishonesty in different conditions. In the control condition, the advisors' payment was linked to the accuracy of the estimator, such that the better the accuracy of the estimator, the higher the advisor's payment. The advisor was therefore motivated to provide the most accurate advice possible, as dishonesty in this condition would be self-harming and lead to lower payouts. The experimenters observed high levels of honesty in this condition (**Fig. 1a**).

To assess self-serving dishonesty, the authors developed two clever scenarios in which the payout for the advisor increased according to how much the estimator overestimated the amount of money in the jar. The advisor was led to believe that the estimator did not know this. Therefore, the advisor had a strong incentive to be dishonest, as he or she could get a larger payout by suggesting larger amounts. In one treatment, dishonesty benefitted both the advisor and the estimator (self- and other-serving), while in another, dishonesty was self-serving and other-harming, such that overestimation increased payment of the advisor at the expense of the estimator. Relative to the control condition, in which dishonesty was negatively incentivized, one would expect greater levels of dishonesty in the two treatment conditions. Moreover, based on previous research[8], dishonesty should be even greater when it is both self- and other-serving, compared to when dishonesty is self-serving and other-harming. Behavioral results from Garrett *et al.*[7] confirmed these predictions (**Fig. 1a**).

The authors also investigated the dynamics of this effect by assessing how dishonesty evolves over a period of 60 trials within these three treatment conditions. When dishonesty was self-harming, dishonesty did not increase significantly over time. In the two conditions, however, in which dishonesty was self-serving, the authors observed a small but gradual increase in dishonesty, reflected by increasing magnitudes of the advisor's recommended overestimation (**Fig. 1b**). These results indicate that dishonesty escalates when it is self-serving and therefore support prior research showing a slippery slope of dishonesty[6].

Does such an escalation occur generally when participants have the opportunity to be dishonest, or is it specific for self-serving dishonesty? To address this question, the authors performed an additional behavioral experiment that included two new treatment conditions, one in which dishonesty was purely self-serving without affecting the payout of the estimator and one in which dishonesty was purely other-serving without affecting the payout of the advisor. Participants' behavior showed an escalation of dishonesty only when it was self-serving but not when it was other-serving, suggesting that selfish but not altruistic motives promote dishonesty adaptation.

How is the slippery slope of self-serving dishonesty represented in the brain? The authors focused their analysis of the fMRI data on the amygdala, a region that is intimately involved in processing emotions, as indicated by meta-analytic results suggesting consistent and relatively specific emotion-related amygdala activity (**Fig. 1c**). The dynamics of neural adaptation to dishonesty were modeled by a variable that assigned decreasing significance to dishonesty as the experiment progressed (**Fig. 1c**). This variable captures the notion that repetition of dishonest acts reduces its emotional saliency and should therefore gradually reduce the neural responses to dishonesty in relevant brain areas. The authors demonstrate that the BOLD signal in the amygdala gradually adapts to dishonesty over time. Note that these results cannot be explained by the amygdala signal simply decreasing over the course of the experiment in the absence of dishonesty nor by the adaptation of the amygdala signal to rewards. The authors took great care to control for such confounding factors by randomly varying the payout amounts on each trial. This prevented

Jan B. Engelmann is at the Center for Research in Experimental Economics and Political Decision Making (CREED), University of Amsterdam, Amsterdam, the Netherlands; and at the Tinbergen Institute, Amsterdam, the Netherlands. Ernst Fehr is in the Department of Economics, University of Zurich, Zurich, Switzerland.
e-mail: j.b.engelmann@uva.nl or ernst.fehr@econ.uzh.ch

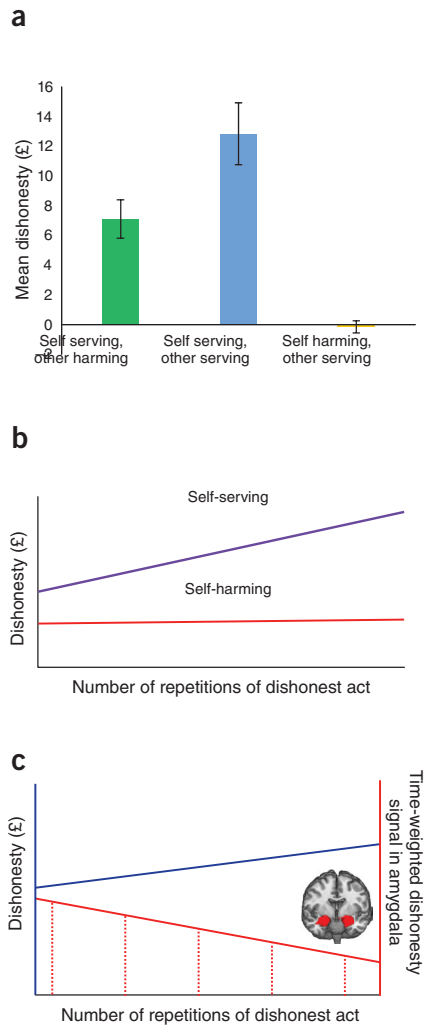

Figure 1 Self-serving dishonesty leads to behavioral and neural adaptation. (**a**) Mean dishonesty levels differed across conditions and were negligible when dishonesty was self-harming, intermediate when dishonesty was self-serving but other-harming and greatest when dishonesty benefitted both the self and the interaction partner (adapted from ref. 7; error bars show s.e.m.). (**b**) Schematic representation of dishonesty escalation. Garrett *et al.*[7] observed a significantly positive slope in the two conditions in which dishonesty was self-serving (purple line) but not when it was self-harming (red line). (**c**) Schematic representation of the neural correlates underlying dishonesty adaptation over time. The upward-sloping blue line illustrates the slippery slope of behavioral dishonesty demonstrated by Garrett *et al.*[7] in the self-serving condition. The vertical dashed red line segments, which are fitted by a solid red line to highlight the downward slope, indicate the decreasing BOLD signal in the amygdala to five instances of dishonesty in which the dishonesty level was identical (for example, the advisor had recommended a £5 overestimate). Despite the identical level of dishonesty in these five instances, amygdala activity was lower when it occurred later in the experiment, when a higher number of dishonest acts had already been committed. The inset shows voxels in the bilateral amygdala (red) that are consistently and relatively selectively associated with emotion as suggested by the reverse-inference map from Neurosynth (an online tool for conducting automated large-scale meta-analyses of fMRI data; http://www.neurosynth.org/).

the subjects from being able to predict the value of lying on a given trial, excluding the possibility of signal adaptation to reward.

Finally, the authors establish an even closer link between trial-by-trial reductions in amygdala activity and increasing levels of dishonesty via prediction analyses. To this end, the amygdala signal on a given trial was first normalized to one monetary unit of dishonesty. The authors then demonstrate that reductions in this normalized signal (from the previous compared to the current trial) predict the escalation of dishonesty (from the current compared to the next trial). These results establish a direct link between reductions in the amygdala signal and the escalation of self-serving dishonesty.

Taken together, the results from Garrett *et al.*[7] are an important first step toward shedding light on the neural mechanisms underlying dishonesty adaptation in the amygdala. However, their provocative findings also raise important new questions. While the decrease in amygdala reactivity to repeated acts of dishonesty is consistent with the hypothesis that one mechanism underlying dishonesty adaptation is a blunted emotional response to the act of lying (for example, the emotional cost when experiencing guilt), the link between emotion and dishonesty adaptation needs to be tightened by future research. This could, for instance, be accomplished by tracking autonomic arousal levels while participants adapt to dishonesty. Blunted emotional responses should be accompanied by decreasing levels of arousal. It is also fundamental to capture the subjective nature of dishonesty adaptation and assess people's awareness of their decreasing moral standards. Using self-reports in combination with psychophysiological measures should enable future research to identify the role of specific emotions in dishonesty adaptation.

Recent meta-analytic accounts of the neurobiological basis of emotion suggest that the brain does not honor distinctions between affective and cognitive mechanisms[9]. Instead, emotions are likely processed in complex networks that involve interactions between 'affective' and 'cognitive' brain regions[10]. It is therefore important for future research to investigate how the amygdala signal interacts with interconnected areas that are involved in neural computations related to dishonesty. Two recent studies[11,12] identify the dorsolateral and dorsomedial prefrontal cortex as important nodes involved in evaluating honesty. One intriguing idea is that inputs from the amygdala to a wider putative dishonesty network, of which the valuation system is a likely constituent[13], are gradually weakened during the process of dishonesty adaptation.

Finally, it is important to take into account individual differences in dishonesty and the efficacy with which a slippery slope can induce increasing levels of dishonesty in different people. One possible explanation for the outrageous levels of dishonesty exhibited by certain individuals may be the formation of habits, in which the amygdala plays a central role[14]. The current results suggest that habitual dishonesty may evolve with greater ease in participants with blunted amygdala reactivity. Pharmacological manipulations during dishonesty adaptation could be used to probe the role of specific neurotransmitter systems in such habit formation. One obvious candidate is the serotonin system, which has recently been implicated in amygdala reactivity to emotional stimuli[15]. Selective enhancements of serotonin neurotransmission to reduce the reactivity of the amygdala during experiments assessing the process of dishonesty adaptation would be expected to diminish the emotions typically triggered by dishonest acts and thereby increase the ease with which people habituate to dishonesty.

The results of Garrett *et al.*[7], together with these considerations, open up exciting venues for future research investigating the emotional mechanisms, extended network interactions and individual differences involved in the slippery slope of dishonesty escalation.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Schwartz, S.H. & Bardi,. A. *J. Cross Cult. Psychol.* **32**, 268–290 (2001).
2. Shalvi, S., Dana, J., Handgraaf, M.J.J. & De Dreu, C.K.W. *Organ. Behav. Hum. Decis. Process.* **115**, 181–190 (2011).
3. Gächter, S. & Schulz, J.F. *Nature* **531**, 496–499 (2016).
4. Mazar, N., Amir, O. & Ariely, D. *J. Mark. Res.* **45**, 633–644 (2008).
5. Cohn, A., Fehr, E. & Maréchal, M.A. *Nature* **516**, 86–89 (2014).
6. Welsh, D.T., Ordóñez, L.D., Snyder, D.G. & Christian, M.S. *J. Appl. Psychol.* **100**, 114–127 (2015).
7. Garrett, N., Lazzaro, S.C., Ariely, D. & Sharot, T. *Nat. Neurosci.* **19**, 1727–1732 (2016).
8. Weisel, O. & Shalvi, S. *Proc. Natl. Acad. Sci. USA* **112**, 10651–10656 (2015).
9. Lindquist, K.A., Wager, T.D., Kober, H., Bliss-Moreau, E. & Barrett, L.F. *Behav. Brain Sci.* **35**, 121–143 (2012).
10. Pessoa, L. & Engelmann, J.B. *Front. Neurosci.* **4**, 17 (2010).
11. Dogan, A. et al. *Sci. Rep.* http://dx.doi.org/10.1038/srep33263 (2016).
12. Baumgartner, T., Fischbacher, U., Feierabend, A., Lutz, K. & Fehr, E. *Neuron* **64**, 756–770 (2009).
13. Bartra, O., McGuire, J.T. & Kable, J.W. *Neuroimage* **76**, 412–427 (2013).
14. Lingawi, N.W. & Balleine, B.W. *J. Neurosci.* **32**, 1073–1081 (2012).
15. Murphy, S.E., Norbury, R., O'Sullivan, U., Cowen, P.J. & Harmer, C.J. *Br. J. Psychiatry* **194**, 535–540 (2009).