Journal of Economic Behavior & Organization 137 (2017) 428–444

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

# The wages of dishonesty: The supply of cheating under high-powered incentives

Parasuram Balasubramanian[a], Victor M. Bennett[b], Lamar Pierce[a,*]

[a] Washington University in St. Louis, United States
[b] Duke University, United States

ARTICLE INFO

*Article history:*
Received 11 October 2016
Received in revised form 16 March 2017
Accepted 27 March 2017
Available online 2 April 2017

*Keywords:*
Incentives
Dishonesty
Cheating
Ethics
Reference points
Prospect theory

ABSTRACT

We use a novel design to identify how dishonesty changes through a broad reward range that, at the high end, exceeds participants' average daily wages. Using a sample of online Indian workers who earn bonuses based on six simultaneous coin flips, we show that the relationship between dishonesty and financial rewards depends on the incentive range. We find two novel effects as incentives exceed those used in most prior research. First, dishonesty increases and reaches its maximum as rewards increase from $0.50 to $3 per reported head and as earnings reach $15, indicating that rewards can indeed motivate more cheating when large enough. More importantly, we show that dishonesty declines at the highest reward levels (up to $5 per head) as individuals appear to engage in lower magnitudes of dishonesty. We detail how our results could be explained by a reference-dependent utility with internal costs of dishonesty that are convex in the magnitude of the lie, and show survey and simulation-based evidence that support this explanation.



## 1. Introduction

One of the most important unanswered questions on dishonesty is how rewards shape decisions to cheat. Although traditional economic models predict that higher rewards from dishonesty should incentivize more cheating (Becker, 1968), more recent work has argued that the internal cost of dishonesty is sufficient to restrain this behavior. Psychologists (Mazar et al., 2008) and some economists (Fischbacher and Föllmi-Heusi, 2013) have argued that since the internal costs of dishonesty from guilt or self-image degradation rise with the magnitude of rewards, increased incentives produce little growth in dishonesty, and may even reduce dishonesty if the internal costs are high enough. This model of increasing internal costs is supported by several studies showing a preponderance of partial liars who restrict the magnitude of their dishonest earnings (Mazar et al., 2008), even with little risk of detection (Shalvi et al., 2012; Fischbacher and Föllmi-Heusi, 2013; Cohn et al., 2014).[1]

* Corresponding author.
*E-mail addresses:* parasuram@wustl.edu (P. Balasubramanian), vmb10@duke.edu (V.M. Bennett), pierce@wustl.edu (L. Pierce).

[1] A closely related literature studies dishonesty using the deception game (Gneezy, 2005), where dishonesty affects the payoffs to another party. This research generally finds that dishonesty increases with financial incentives (Sutter, 2009; Dreber and Johannesson, 2008; Erat and Gneezy, 2012), although the third study in recent work by Wang and Murnighan (2016) shows no consistent relationship between incentives and lying. Related studies that examine how incentives impact selfishness similarly suggest that the range of incentives matters. Although early work typically finds financial stakes to have little

http://dx.doi.org/10.1016/j.jebo.2017.03.022
0167-2681/

Indeed, laboratory studies where participants can dishonestly increase their earnings have found no identifiable relationship between incentives and lying. Mazar et al. (2008) found no difference in self-reported matrix-task performance between two incentive conditions ($0.50/matrix and $2/matrix) in Study 2. They also report that an additional study found no identifiable cheating when incentives were extended to $5/matrix, although they do not report enough details to evaluate the incentive-dishonesty relationship. Similarly, Fischbacher and Föllmi-Heusi (2013) found no increase in dishonesty among students reporting a secret die roll when maximum rewards tripled from 5 to 15 Swiss Francs (approximately $5–$15).

In this paper, we argue that although these studies provide strong evidence for a low range of incentives, they do not allow for a more complex relationship between incentives and dishonesty that could be non-monotonic over a broader range of financial rewards. Prior work has used a limited number of incentive conditions, mapping a linear relationship within that range. Furthermore, the value range of these conditions is relatively low considering the high wealth levels of their Western participant populations, so we know little about dishonesty in higher incentive ranges. Mazar et al. (2008) use the highest incentives, with maximum earnings of $20 without substantial risk of detection.[2] Understanding how larger reward ranges shape dishonesty is particularly crucial because they more closely parallel organizational and other field settings, thereby improving the generalizability of this research stream. Managers, policy-makers, and societies primarily focus on dishonest behaviors that yield much larger rewards than existing studies represent.

Our focus on a broader incentive range is also motivated by several recent studies that suggest that cheating indeed occurs at high reward levels. Cohn et al. (2014), for example, found substantial cheating when using a single incentive level where participants could earn $200, although their study invoked competitive factors that are known to accelerate cheating (Bennett et al., 2013; Branco and Villas-Boas, 2015; Kilduff et al., 2016). Weisel and Shalvi (2015) also found more cheating in the higher of two incentive levels in a design that required collaborative deceit. Evidence from Swedish tax returns also shows increased dishonesty at higher reward levels (Engström et al., 2015).

More closely related to our work is Kajackaite and Gneezy (2017), who use four self-reported die-roll conditions with binary payoffs (rolling a 5 earns $1, $5, $20, or $50). Although they indeed find a non-monotonic relationship between rewards and dishonesty—cheating is highest at $20 and negligible at $50—they attribute the decreased dishonesty in the highest condition to the threat of detection. Four additional conditions, where participants self-reported whether their die roll matched a previously imagined number, showed increased dishonesty at the highest incentive levels. Although these "mind game" conditions could indeed reflect the impossibility of lie detection, they also could be explained by motivated forgetting (Shu et al., 2011) or the justification effect observed in Shalvi et al. (2011)—both of which would predict higher cheating levels.

We attempt to map and understand the relationship between rewards and dishonesty through two studies that manipulate a wide range of incentive levels in an online labor market with lower wealth levels than the standard U.S. and European experimental populations. In the first study, we employ Indian workers on Amazon's Mechanical Turk (MTurk) platform to complete common image recognition tasks, then pay them bonuses based on self-reported outcomes from six simultaneous coin flips on a third-party website (random.org). Our experimental design has three major advantages over prior work mapping rewards to dishonesty. First, our sample of Indian workers allows us to manipulate bonus rates in ten conditions from $0.50/head to $5/head, such that in the top condition workers who report 6 heads earn more than their average daily wage in a short period of time. The magnitude of the incentives, in terms of purchasing power, in our highest conditions is thereby similar in magnitude to the top $50 condition used in a concurrent working paper by Kajackaite and Gneezy (2017) with University of California students. Second, our use of six coin flips instead of the six-sided die in many prior studies allows us to detect dishonesty with fewer participants because of the binomial distribution's lower probability of honestly achieving an extreme outcome (e.g., six heads). Third, our use of ten conditions allows us to better map any non-linear relationship between rewards and dishonesty.

We find that although dishonesty is identifiable at every reward level, it is highest in the mid-level conditions of $2.50/head and $3/head with strong evidence of non-monotonicity. Reported head counts are lower in both the highest and lowest conditions due to decreases in both lying magnitude and frequency. Although we cannot directly measure the mechanism driving this non-monotonic relationship between rewards and dishonesty, we show that it could result from the reference-dependent utility associated with prospect theory (Kahneman and Tversky, 1979; Barberis, 2013), where reference points related to expected daily income might influence marginal decisions to cheat. Just as unexpectedly high earnings might lead to decreased effort (Köszegi and Rabin, 2006; Abeler et al., 2011) in taxi drivers (Camerer et al., 1997) or bike messengers (Fehr and Goette, 2007), so too might unexpectedly high earnings from both honesty and cheating (Dugar and Bhattacharya, 2017) allow individuals to curtail costly dishonesty. Recent evidence (Kern and Chugh, 2009; Engström et al., 2015; Grolleau et al., 2016; Rees-Jones, 2017; Garbarino et al., 2016) indeed shows that losses might motivate cheating more than gains, but this effect has not been mapped to income reference points. Also related is evidence that goals, which can function as reference points (Heath et al., 1999), can shift cheating decisions (Schweitzer et al., 2004).

impact on selfish behavior in experiments (Camerer and Hogarth, 1999; Cherry et al., 2002), later work using higher incentive ranges suggests decreased transfers (Carpenter et al., 2005) and costly punishments (Andersen et al., 2011) as the stakes reach economically meaningful levels for participants.

[2] Although their task provides maximum earnings of $40 by reporting 20 successful matrices in 4 minutes, 10 of the matrices are unsolvable. This likely applies an additional constraint on the magnitude of lying—the threat of detection—even in their condition using a shredder.

We conduct a second study that explores the reference-dependent explanation through a follow-up survey of Study 1 participants, gathering information about daily income, target income, and daily stopping decisions. These data indicate likely reference points that are consistent with prospect theory utility functions with convex costs of dishonesty, and suggest loss aversion as a possible mechanism for our Study 1 results. Although the income reference points from these data are speculative, unlike more objective references such as the round number times for marathon runners (Allen et al., 2016), the self-reported data are consistent with the range of reference points that would produce our data under reference-dependent utility. A third study addresses participant concerns of lie detection and risk of non-payment.

## 2. Study 1

### 2.1. Experiment design

For our participant pool we recruited Indian workers from MTurk (Goodman et al., 2013; Litman et al., 2015) using the TurkPrime interface (Litman et al., 2016). Although Indian workers used to represent about one-third of total workers on MTurk (Ipeirotis, 2010), changes in account requirements are estimated to have decreased this share to 6–8% (Goodman, 2015). In both time periods, most Indian workers report using MTurk as their primary source of income. We limited our study to Indian workers to help generate higher incentive conditions relative to daily wages, at a relatively low cost. While the Indian economy continues to grow quickly, the national price level index is estimated at 3.6x (U.S.), 3.8x (Germany), and 5.4x (Switzerland)[3]—much lower than countries where related studies were conducted. All experimental procedures were approved by the principal investigator's Institutional Review Board.

We initially invited 1000 workers to complete a demographic survey (see Appendix) for a fixed fee of $0.50. This survey served two primary purposes. First, it allowed us to gather detailed information on participant gender, location, religion, and other demographics at a time point different from our primary study. Second, it allowed us to prescreen potential participants to ensure demographic diversity, vet potentially fraudulent repeat participants, and reduce self-selection concerns based on our substantially higher-than-average earnings. Of the 998 respondents, 907 remained after deleting workers who appeared to use the same IP address (possibly multiple accounts) or did not complete the survey. Because many workers use dynamic IP addresses, our initial deletion could still miss some workers who completed the survey more than once. We then used an additional level of screening by dropping duplicate responses across six criteria – year of birth, gender, marriage status, state, education and religion. The probability that two or more workers will randomly match all six criteria together is low if they were truly independent respondents. This step decreased our sample to 660 responses.

We then used worker IDs collected from the demographic survey to invite a stratified random subsample of 397 (out of the 660) workers to participate in a second work task. We used a two-step process for creating our sample. First, because the vast preponderance of our respondents were from Tamil Nadu and Kerala, to assure geographic diversity, we included all respondents from states besides those two. Second, within Tamil Nadu and Kerala, strata were created to ensure variation in gender, age, and income, assigning each observation the product of the probability of appearing in the sample if the population was uniformly distributed across age classes by the probability of appearing in the demographic survey. For example, 166 of the 660 candidates were in the lowest income category of seven, annual household income <100k INR. Observations in this category, therefore, had their probability of being invited by $\frac{\frac{1}{7}}{\frac{166}{660}}$. 331 (83.4%) of the 397 invited workers started the second work task, of which 320 completed it.

Table 1 shows demographic variation for our initial sample of 907 workers who completed the demographic survey, 660 workers whom we accepted as candidates for the bonus task based on surviving the screening for duplicate participants, the 397 workers who were invited to participate in the bonus task, and the 320 workers who completed it. Our screening criteria and strata focused on geographic diversity resulted in little variation of worker demographics across the three samples. The median category for each variable is highlighted in the table below. For all categories the median remains the same across samples–the median worker is between 26 and 35, has a household income of between 200,001 INR and 500,000 INR ($8500), has a college degree, is married and spends 8–20 hours weekly on MTurk. Of the total 320 responses, 35% of the sample are women, consistent with the aggregate percentage of female MTurk workers in India.

With a diverse country such as India, it is a challenge to get a truly representative sample of regional variation. Fig. 1 shows our sample representation of 24 states and union (federal) territories, with the states of Karnataka, Kerala, Maharashtra and Tamil Nadu contributing over 50% of responses.

The second task used the Qualtrics platform to deliver 30 image recognition tasks that paid $0.05 per correct response. The estimated time to complete this task was around 15 minutes, resulting in an hourly wage of $6 per hour. The labor task, which involved identifying restaurant names from signs in two Midwestern U.S. cities (see Appendix), was intentionally easy in order to avoid differential performance outcomes between workers. Since the image recognition task involved text responses, we anticipated that there could be spelling errors or differences in use of fonts (e.g. all caps or all lower case). We

[3] World Bank, International Comparison Program database. http://data.worldbank.org/indicator/PA.NUS.PPPC.RF. This ratio obtained by dividing purchasing power parity (PPP) conversion factor by the market exchange rate makes it possible to compare the cost of a bundle of goods that makes up a country's GDP. One can compare how many dollars are required to buy a dollar's worth of goods in a country as compared to the U.S.

**Table 1**
Demographics for Each Stage of Sample Construction.

| Demographic variation across samples | 907 obs. | 660 obs. | 397 obs. | 320 obs. |
|---|---|---|---|---|
| % Female | 34.2% | 34.9% | 37.8% | 35.0% |
| Age | | | | |
| 18–25 | 21.1% | 18.5% | 19.6% | 18.8% |
| 26–35 | 53.9% | 52.7% | 52.9% | 52.5% |
| 36–45 | 18.0% | 19.4% | 18.1% | 19.1% |
| 46–60 | 5.6% | 7.4% | 7.3% | 7.5% |
| Over 60 | 1.4% | 2.0% | 2.0% | 2.2% |
| Annual Household Income (in INR) | | | | |
| Less than 100,000 | 17.2% | 15.9% | 16.9% | 15.3% |
| 100,001–200,000 | 25.8% | 25.2% | 19.6% | 19.7% |
| 200,001–500,000 | 35.0% | 34.5% | 32.5% | 32.5% |
| 500,001 to 1 Mn. | 16.6% | 17.4% | 19.4% | 20.6% |
| More than 1 Mn. and up to 2.5 Mn. | 4.6% | 5.9% | 9.8% | 9.7% |
| More than 2.5 Mn. and up to 5 Mn. | 0.4% | 0.6% | 1.0% | 1.3% |
| More than 5 Mn. | 0.3% | 0.5% | 0.8% | 0.9% |
| Education | | | | |
| High School/Diploma | 5.2% | 7.0% | 5.5% | 4.4% |
| Bachelors Degree | 55.2% | 52.3% | 52.4% | 52.8% |
| Masters Degree | 35.9% | 35.9% | 36.3% | 36.6% |
| Professional Degree (JD, MD) | 2.0% | 2.7% | 3.5% | 3.8% |
| Doctoral Degree | 1.4% | 1.8% | 2.0% | 2.5% |
| Other | 0.2% | 0.3% | 0.3% | 0.0% |
| Marriage Status | | | | |
| Married | 61.0% | 62.7% | 62.7% | 62.5% |
| Unmarried | 36.5% | 33.8% | 33.2% | 34.1% |
| Other | 2.5% | 3.5% | 4.1% | 3.4% |
| Hours per Week on MTurk | | | | |
| Less than 1 h | 2.8% | 2.6% | 2.8% | 2.2% |
| 1–2 h | 8.4% | 8.9% | 9.8% | 8.8% |
| 2–4 h | 15.1% | 14.7% | 13.4% | 13.4% |
| 4–8 h | 18.5% | 19.8% | 21.4% | 19.1% |
| 8–20 h | 24.1% | 24.1% | 25.9% | 26.6% |
| 20–40 h | 16.4% | 15.0% | 14.1% | 15.9% |
| More than 40 h | 14.6% | 14.7% | 12.6% | 14.1% |

Notes: Demographic variation for initial, post-screening, invited workers and final sample for the study. Shaded observations represent the median category for each variable which remained the same during study 1. 907 workers completed the initial demographic survey, 660 workers were accepted after screening for duplicates, and 397 workers were invited for the bonus task of which 320 workers completed the bonus task in Study 1. Indian workers in our study were largely under the age of 45 and had a college degree, while the gender ratio remained nearly the same in study 1. See table A1 in Appendix for demographic details by gender for the bonus task.

were thus lenient with minor differences and approved a worker so long as she attempted to provide the restaurant name. For example, if the name in the image was "TED DREWES", we counted as correct answers that included "ted drewes", "Ted Drewes", or even the misspelling "Ted Dreues". We verified legitimate effort manually for each worker before approving them for payment. Our final sample of 320 workers thus contains only workers who legitimately attempted all 30 image recognition tasks. Seven workers who chose to skip through the images to the bonus task were rejected.[4]

Following completion of the image recognition task, workers were randomly assigned into one of ten conditions that offered monetary bonuses of different expected values. Workers were instructed to visit a third-party website (random.org), and told to click once on "Flip Again" on a page that showed six British £1 Sterling coins (see Fig. 2) with randomly determined outcomes. Upon flipping the six coins simultaneously (which produced a new random result), workers were asked to report the number of heads in Qualtrics. Based on the condition, workers were promised and given a bonus ranging from $0.50 per head to $5 per head reported (in $0.50 increments), for potential earnings between $0 and $30. Workers did not know the payoffs of any conditions other than their own, which was important to avoid envy- or inequity-based concerns identified in prior work (Gino and Pierce, 2009; Gino and Pierce, 2010; John et al., 2014). Although it is possible that workers' reporting decisions after clicking "Flip Again" were influenced by the existing head count when arriving at the page, the randomness of these existing head counts ensures that such an effect would be constant across conditions. Because we used a third-party website, monitoring the true outcome of any given coin-flip set was impossible. Workers could therefore dishonestly over-report any outcome without risk of detection. Furthermore, even the highest-paying outcome (six heads) was plausible, given that in expectation some workers would indeed reach that outcome. Dishonesty could only be observed in aggregate

[4] This includes 1 worker each in $1.50, $3.50 and $5.00 conditions, 2 workers each in $2.50 and $4.50 conditions; in addition 3 workers did not complete the image recognition task nor proceed to the coin flipping task, and 1 worker who did not complete the coin flipping task.

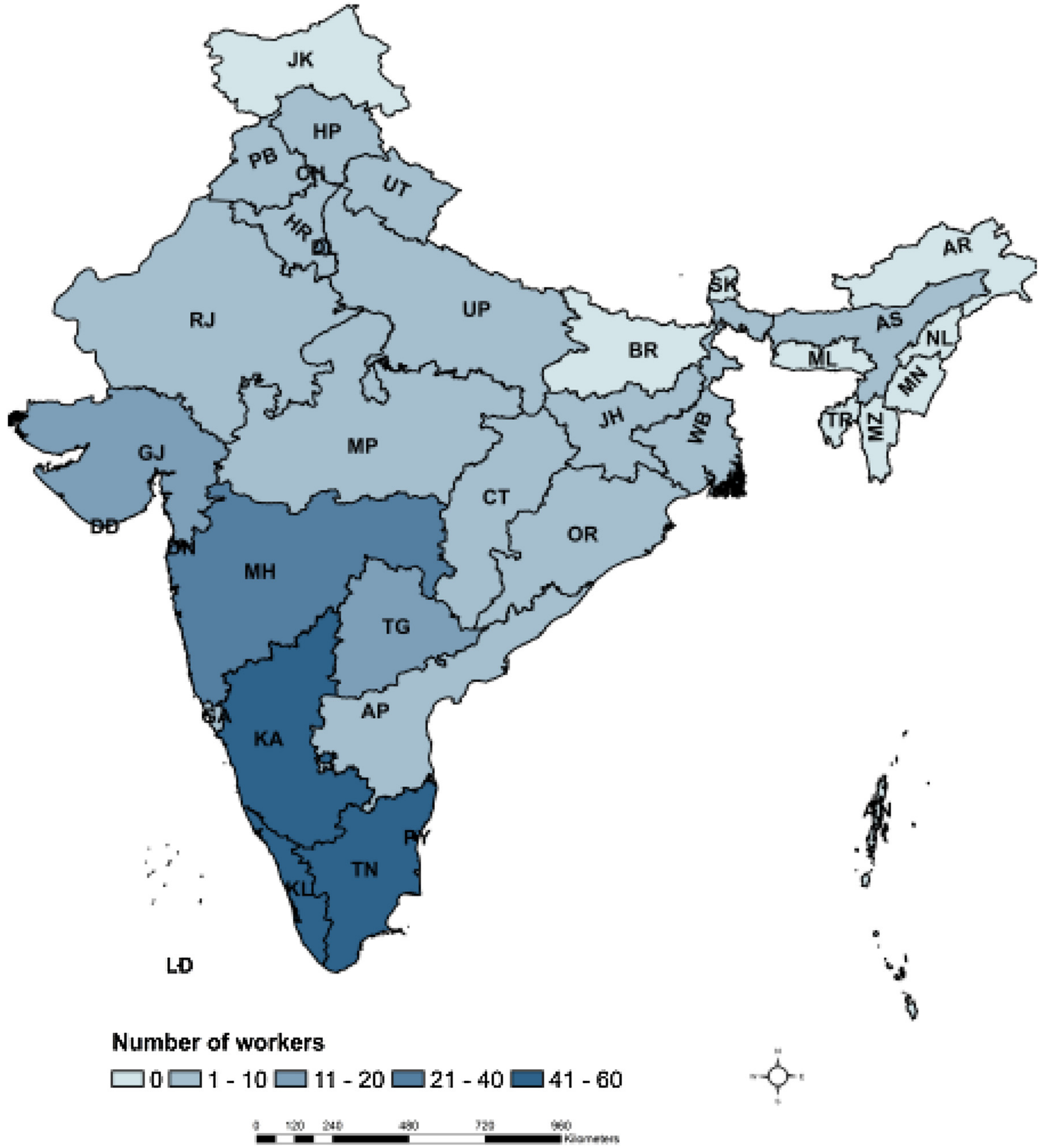


**Fig. 1.** Geographic Distribution of 320 Full Participants.
*Notes:* India map with distribution of workers by state and union (federal) territory from http://www.gadm.org. Populations are in parentheses. AP = Andhra Pradesh(49.5 M), AR = Arunachal Pradesh(1.3 M), AS = Assam(31.2 M), BR = Bihar(103.8 M), CT = Chhattisgarh(25.5 M), GA = Goa(1.5 M), GJ = Gujarat(60.4 M), HR = Haryana(25.4 M), HP = Himachal Pradesh(6.9 M), JK = Jammu and Kashmir(12.6 M), JH = Jharkhand(33.0 M), KA = Karnataka(61.1 M), KL = Kerala(33.4 M), MP = Madhya Pradesh(72.6 M), MH = Maharashtra(112.4 M), MN = Manipur(2.7 M), ML = Meghalaya(3.0 M), MZ = Mizoram(1.1 M), NL = Nagaland(2.0 M), OR = Odisha(41.9 M), PB = Punjab(27.7 M), RJ = Rajasthan(68.6 M), SK = Sikkim(0.6 M), TN = Tamil Nadu(72.1 M), TG = Telangana(35.2 M), TR = Troipura(3.7 M), UP = Uttar Pradesh(199.6 M), UT = Uttarakhand(10.1 M), WB = West Bengal(91.3 M).

by comparing distributions of coin flip outcomes with the known counterfactual distribution of honestly-reported fair-coin outcomes.

Our use of multiple independent coin flips, similar to Abeler et al. (2014), Cohn et al. (2014), and Peer et al. (2014), provides a key advantage over the more standard die-roll task. Since coin flips follow the binomial distribution, the head-count of six fair coins produces a normal distribution whose highest outcomes (5 or 6) occur with greater rarity than with the uniformly distributed die roll (see Fig. 3). As Figs. 4 and 5 show, this provides tighter confidence intervals and allows the researcher to identify equivalent levels of cheating with fewer observations than in the die roll task.


**Fig. 2.** Screenshot of Bonus Task Webpage.
*Notes:* Screenshot of random.org page to which workers were directed and told to click "Flip Again". The initial head count upon accessing the website is random, so each worker saw a different combination of heads and tails.

Our bonus structure based on the number of heads (as opposed to a die roll) also allows us to partially separate the average magnitude of cheating from the expected incidence of cheating. which is important in establishing the internal cost structure of dishonesty. Evidence from both psychology and economics shows that many individuals choose to partially lie, which suggests that dishonesty costs are not purely fixed but rather dependent on lying magnitude (Mazar et al., 2008; Fischbacher and Föllmi-Heusi, 2013). We are able to separately calculate changes across condition in both the likelihood and magnitude of cheating by exploiting the non-uniform probability distribution of our multiple coin-flip task (see Appendix for details).

### *2.2. Results*

Fig. 5 presents the mean number of reported heads with 95% confidence intervals for each incentive condition (see Appendix for count distributions by condition). Confidence intervals were calculated by simulating 600 fair coin flips (100 individuals) per condition. We observe identifiable cheating in each condition; each mean is statistically distinguishable from 3 ($p < 0.05$). This contradicts evidence from many prior studies (Mazar et al., 2008) that cheating disappears at higher incentive levels. Furthermore, cheating levels are lower in the four highest ($3.50–$5) and four lowest ($0.50–$2) conditions,

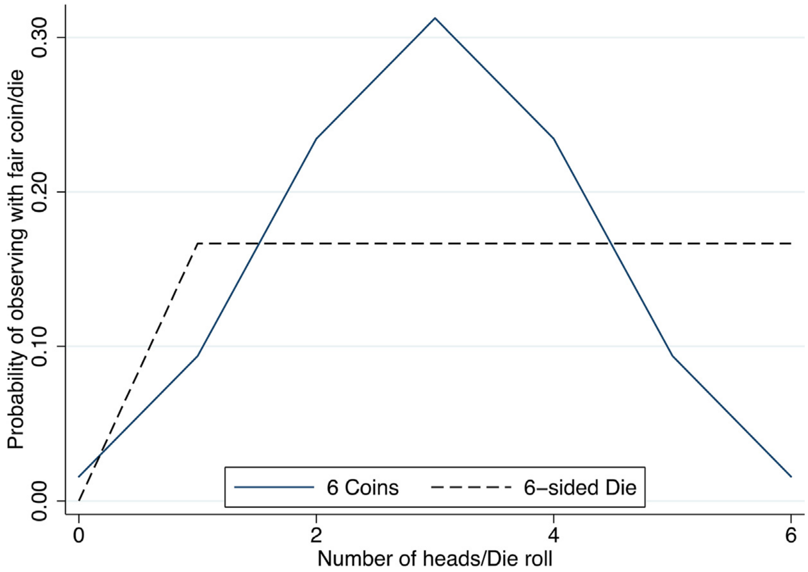


**Fig. 3.** Probability Distributions of Honest Outcomes from Six Coin Flips and a Six-Sided Die.
*Notes:* This figure shows the true probability of observing each possible outcome from alternatively flipping six fair coins or rolling a fair six-sided die. High outcomes (5 or 6) are less likely with coin flips, making dishonest reporting more identifiable in aggregate samples.

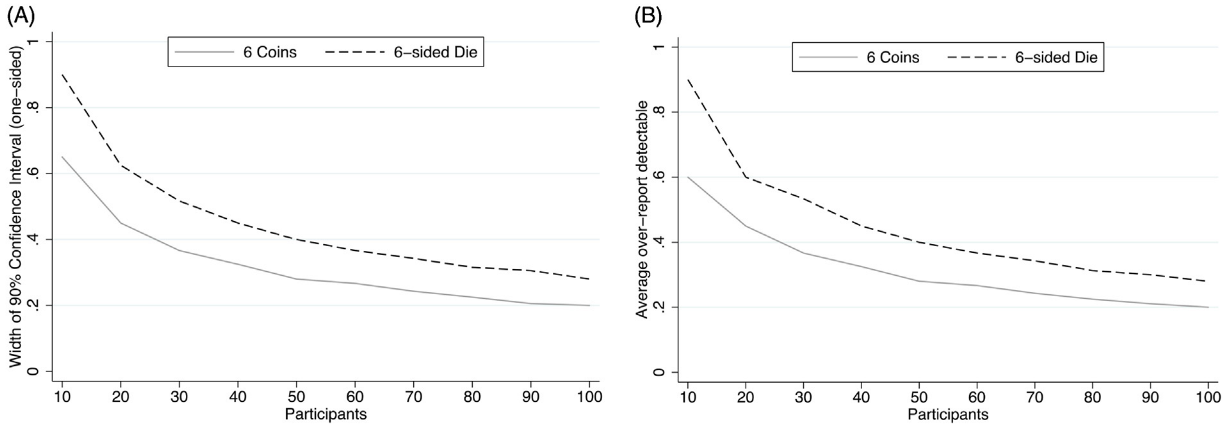


**Fig. 4.** Improved Statistical Power from Coin-Flip Task.
*Notes:* Panel A shows the width of the 90% confidence interval for rejecting honest reporting of either a six-sided die roll or six independent coin flips, as a function of sample size. Panel B shows the minimum average level of over-reporting detectable with 95% confidence as a function of sample size for both coin flips and die rolls.

and substantially higher in the two middle ($2.50 and $3) conditions, suggesting a non-monotonic relationship. Fig. 6 shows the condition means with a lowess fit illustrating the non-parametric relationship.

We formally test for non-monotonicity using a series of statistical procedures. First, we test for a simple linear relationship between rewards and cheating, and find a linear trend that is indistinguishable from zero ($\beta_{condition}=0.067$, $p=0.221$) (see Appendix). We next test whether the data indicate a positive relationship until the maximum ($3) and then a negative relationship above the maximum. We do so using piecewise linear regressions. We first estimate the linear relationship from 0.5 to 2.5 and a separate linear relationship from 3 to 5. The regression results show a negative and statistically significant ($\beta_{condition*high}=-0.467$, $p=0.036$) relationship for the top five conditions and a positive yet imprecise slope for the lower five conditions ($\beta_{condition}=0.191$, $p=0.259$). This imprecision for the lower-range segment results purely from the discrete nature of the independent variable, however, since the maximum value observations ($3) cannot be included in both line segments. Furthermore, the positive slope in the lower conditions is significant and larger if the $3 condition is included in the lower segment ($\beta_{condition}=0.341$, $p=0.006$), while the upper condition regression segment becomes less precise ($\beta_{condition*high}=-0.341$, $p=0.143$). The maximum dishonesty condition, $3, clearly has different slopes on each side.

A non-parametric Wilcoxon-Mann-Whitney test shows that the distribution of reported headcounts for the lowest ($.5) reward condition is lower than headcounts in the peak dishonesty condition ($3) ($p=0.041$). Similarly, the highest reward condition ($5) has lower headcounts than the peak condition ($p=0.034$). The joint probability that the relationship between

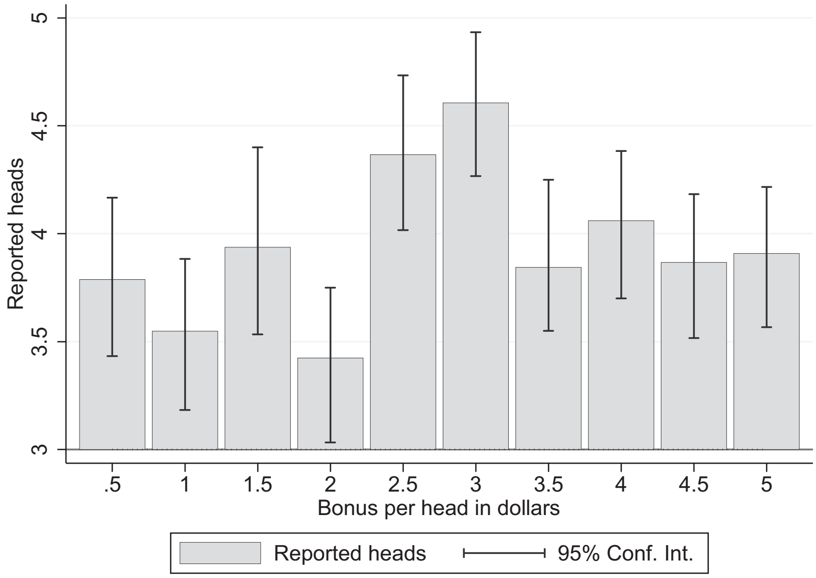


**Fig. 5.** Mean Reported Heads by Condition.
*Notes:* This figure shows the mean reported heads by condition, with 95% confidence intervals generated through simulations using the binomial distribution of a fair coin. The true expected mean for each condition is 3.

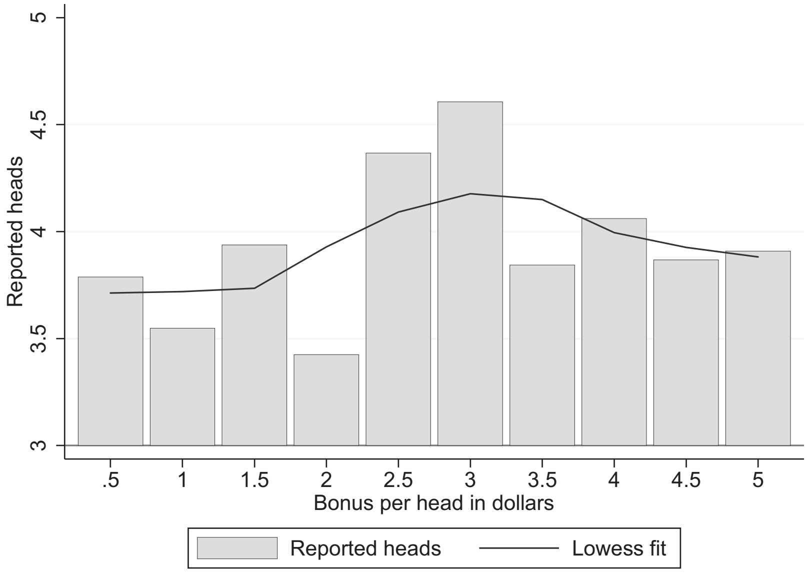


**Fig. 6.** Mean Reported Heads by Condition with Fitted Lowess.
*Notes:* This figure shows the mean reported heads by reward condition, and the line represents a non-parametric lowess. The true expected mean for each condition is 3.

rewards and dishonesty increases to $3 then decreases through $5 is therefore 0.93. We note that these tests are the closest equivalent to the tests conducted in prior studies with only 3–4 reward conditions.

We also conduct non-parametric trend tests (Cuzick, 1985) for the lower ($.5–$3) and upper ($3–$5) reward ranges. The lower range demonstrates a positive trend ($p = 0.051$) while the upper range has a negative trend ($p = 0.008$). We also note that all six (versus expected total number of 5 with honesty) participants who reported zero heads were in the five lowest incentive conditions (see Appendix).

As a final test, we use a quadratic regression combined with the post estimation test for an inverted u-shaped relationship derived by Lind and Mehlum (2010) and provided in Stata through the command *utest*. The u-test weakly supports an inverted u-shape with a lower-bound slope of 0.376 and upper-bound slope of −0.243 ($p = 0.109$). Although we cannot be certain about the parametric relationship between rewards and dishonesty, quadratic, cubic, and quartic regression provide reasonable if imprecise fits (see Appendix for figures).

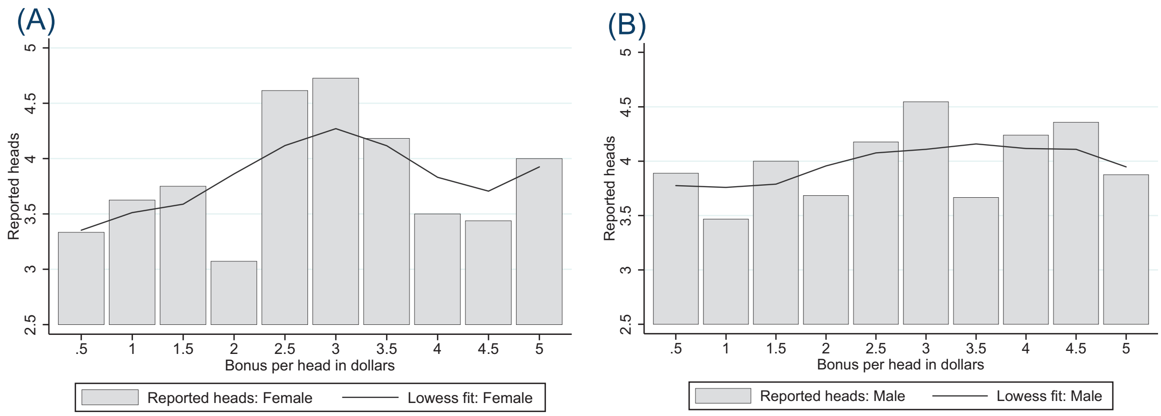


**Fig. 7.** Mean Reported Heads by Gender with Fitted Lowess.
*Notes:* Smoothed lowess fit of reported heads and condition, by gender (Image (A): female; Image (B): male). Women represent about 35% of the sample in the bonus task. While there seems to be more variation in reporting of heads across conditions in females as compared to males, there is no statistically significant difference in cheating across gender. The true expected mean for each condition is 3.

### 2.3. Demographic and cultural variation

We also examined whether the demographic and cultural variation explains variation in cheating in our data. We first explored whether regional corruption levels predict the likelihood of dishonesty among our workers, since prior work (e.g., Mazar and Aggarwal, 2011; Lowes et al., 2017; Gächter and Schulz, 2016) has found modern and historical institutions and culture to explain propensity to cheat. Using regional corruption categories (moderate, high, very high and alarming) from the 2008 India Corruption Study (Transparency International India), we regressed reported head count on regional corruption without controls ($\beta_{corruption} = -0.0407$, $p = 0.629$) and also controlling for age, income, education and gender (Appendix). Unlike prior work, we observe no identifiable correlation between local corruption and cheating, although we note that corruption in all regions of India is at least moderately high.

We also examined whether age, education level, income, and gender play any role in the estimated levels of dishonesty. The results of the regressions with income ($\beta_{income} = 0.139$, $p = 0.230$), worker age ($\beta_{age} = -0.003$, $p = 0.721$), and education ($\beta_{education} = 0.033$, $p = 0.754$), did not have a statistically significant effect on reported headcount in our study (see Appendix). Fig. 7 illustrates the lowess fit of reported heads and condition varied by gender, and exhibits a similar pattern in reporting across conditions.

### 2.4. Cheating rate versus magnitude

Theoretical models of dishonesty raise the question of whether the sharp increase at the mid-level incentives is due to increased rates of cheating (number of cheaters) or increased levels of cheating. Fig. 8 presents the number of low (0–1), mid (2–4), and high (5–6) reported head counts in each condition. The figure shows that a substantial increase in the number of reported fives and sixes explains the mean increase in the middle conditions (see Appendix for full distributions), which suggests that some of the variation in mean head count across conditions is explained by the changes in the magnitude of cheating.

Fig. 9 overlays the frequency (expected proportion of dishonest workers) and magnitude (the expected heads over-reported) of dishonesty by condition. The frequency, measured as the proportion of dishonest workers, is based on the conditional probability calculations presented in Table 2. The expected heads over-reported are calculated as follows: first, we assume the true value was weakly lower than the reported value that no one lied about having a smaller value than actual. Then, for each possible smaller true value we multiply the magnitude of lie necessary to reach the observed value by the conditional probability of that smaller true value (from Table 2).[6] Finally, we sum the expected magnitudes of the lie for each candidate true value. The figure suggests that the non-monotonic relationship between rewards and dishonesty is a function both of choices to lie and choices of *how much* to lie. This is consistent with models of internal dishonesty costs that are magnitude-dependent, not those with fixed costs of cheating.

### 2.5. Possible theoretical explanations for non-monotonicity

What theoretical mechanisms might explain the non-monotonic relationship between rewards and dishonesty? As we noted in the introduction, this relationship is a function both of the benefits of the dishonest gains and costs of the dishonesty

[6] For example, the expected magnitude of the lie for an observed report of 2 heads is 2*(0.0455)+1*(0.2727)+0*(.6818)=0.36.

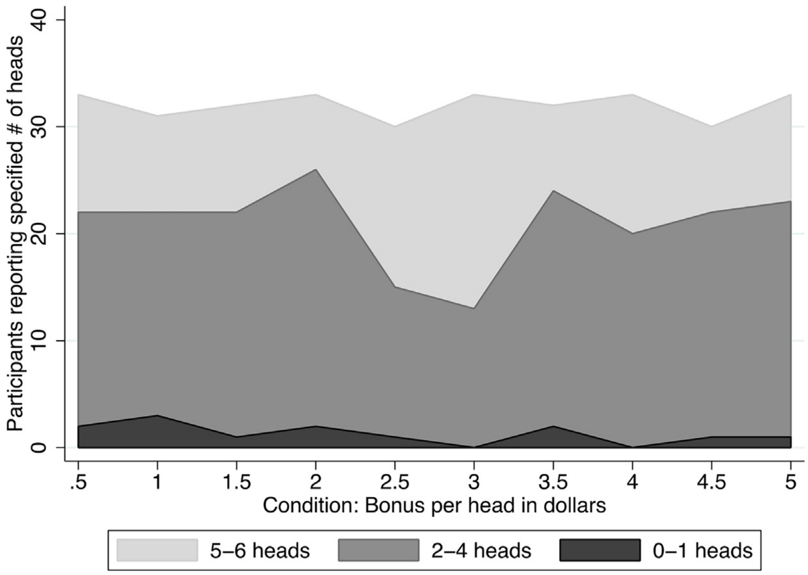


**Fig. 8.** Distribution of Reported Heads by Condition.
*Notes:* This figure shows the reporting of low (0–1), medium (2–4), and high (5–6) outcomes under each reward condition for all workers. We observe an increase in reported 5's and 6's in the middle conditions ($2.5, $3.0), and decreased cheating level under highest incentives. See Appendix for distribution of heads reported by condition.

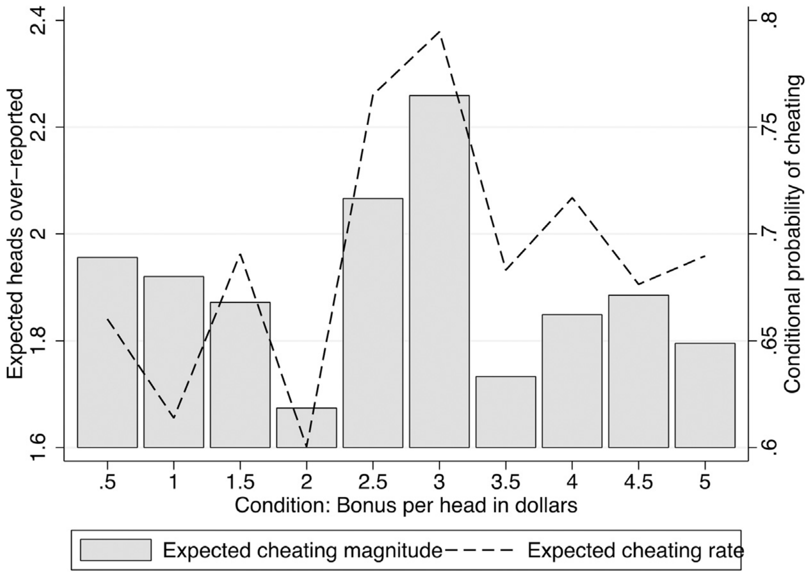


**Fig. 9.** Expected Rates and Magnitudes of Cheating by Condition.
*Notes:* This figure shows the expected cheating rates (dashed line) overlaid on the expected magnitude of heads over-reported (dishonesty) by condition. Expected cheating rates are estimated from conditional probability of cheating, calculated in Table 2. We observe that expected cheating rates decrease in the highest-reward conditions, but remain higher than in the lowest-reward conditions.

necessary to achieve them. Given that our study design has made detectability (and thus punishment) impossible, the relevant costs of dishonesty to this decision must be internal. Prior work suggests that internal costs of dishonesty are either fixed (independent of magnitude) or increasing and even convex in magnitude (Mazar et al., 2008; Kartik, 2009). We do not observe the high proportion of reported 6′s that a model with fixed costs of lying would predict (see Appendix). Below, we use utility functions to examine possible theoretical explanations.

We investigate three candidate utility functions and whether they might yield the observed non-monotonicity with reasonable parameters. For each utility function described below, we do the following. First, conditional on the *b* parameter that defines the convexity of the cost of dishonesty and the value function for money, we numerically solve for the optimal number of reported heads for each number of true heads an individual may have randomly flipped. Second, we use the

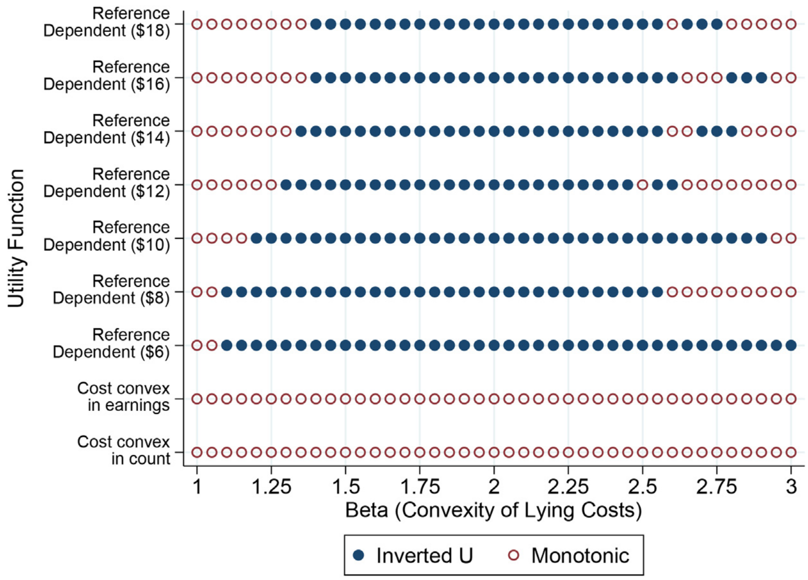


**Fig. 10.** Utility Functions That Produce Non-Monotonic Inverted-U Relationships Between Rewards and Dishonesty.
*Notes:* This figure shows results of the three utility functions shown below, with plausible range of values for convex costs $b > 1$. The reference-dependent utility function is plotted for reference values from $6 to $18 per day. While we do not observe a non-monotonic relationship for the first two utility functions, we observe a non-monotonic inverted-U relationship between reported heads and reward conditions, when $b$ takes values between 1.3 and 2.6, and reference income points range from $6 to $18 per day.

limit probabilities of individuals flipping a number of true heads, and multiply these probabilities by the optimal report to get a limit average number of reported heads for the utility function-condition-parameter value trio. Third, for each pair of value function and $b$, we indicate whether the average reported heads would be non-monotonic in condition, i.e. whether reports have an inverted u-shape in conditions as indicated by the highest and lowest condition reported head counts being less than the peak condition. Fig. 10 presents the results for the three utility functions for a range of plausible values of the convexity of cost $b > 1$, described below, with different reference point values for the reference-dependent function.

### 2.5.1. *Concave benefits and costs of dishonesty convex in number of over-reported heads*

A standard utility function where utility is concavely increasing in lifetime income and convex in costs cannot produce the non-monotonic relationship observed in Study 1. Such a utility function:

$$v(x) = (wx)^a - (x - h)^b \tag{1}$$

where $w$ is the reward per head (condition), $x$ is the reporting choice, and h is the random true head count. As Fehr and Goette (2007) note, Rabin (2000) demonstrates that a standard economic utility model with concave benefits from long-term income is linear in small amounts. This implies that in an experimental design such as ours, standard utility functions with convex costs of dishonesty would take the form:

$$v(x) = \mathrm{wx} - (x - h)^b \tag{2}$$

Although such an equation might limit the magnitude of lying, the optimal choice $x^*$ is always increasing in $w$. Fig. 10 shows that this utility function doesn't yield a negative quadratic response function for any of the tested values of $b$.

### 2.5.2. *Concave benefits and costs of dishonesty convex in gains from over-reporting*

Making the costs of dishonest dependent on the monetary value of the dishonesty instead of the magnitude of the lie also cannot generate our non-monotonic results. This utility is presented in Eq. (3)

$$v(x) = (wx)^a - ((x - h)w)^b \tag{3}$$

where $a < 1$ and $b > 1$. Fig. 10 demonstrates that such a utility function also cannot not generate the observed non-monotonicity for any of the tested values of $b$.

### 2.5.3. Reference-dependent with convex costs of dishonesty

An alternative model assumes that individuals have reference-dependent utility functions with convex costs of dishonesty. Using the reference-dependent utility function from Tversky and Kahneman (1992) with additively separable convex costs of lying, utility is:

$$\begin{aligned} v(x) &= (wx - r)^a - (x - h)^b \text{ if } wx \geq r \\ v(x) &= -\lambda(r - wx)^a - (x - h)^b \text{ if } wx < r \end{aligned} \tag{4}$$

Unlike the other two utility functions, such a model can indeed produce the non-monotonic relationship found in Study 1. We assume a = 0.88 and lambda = 2.25 (as in Tversky and Kahneman (1992)), and vary the convexity of the cost of dishonesty as well as the reference income. Fig. 10 shows that unlike the previous models, such a reference-dependent utility function can produce a non-monotonic inverted-U relationship between reported headcount and rewards for a reasonable range of $b$ between 1.3 and 2.6 and reference points from \$6 to \$18.

Fig. 11A–C demonstrate how a reference-dependent utility function with convex costs of dishonesty can produce results similar to Study 1 using 10,000 simulated participants randomly assigned coin flips and reward conditions. Fig. 11A presents a model with a reference point of 8; Fig. 11B uses a reference point of 12; Fig. 11C uses a reference point of 16. For each of these three reference points, there exists a non-monotonic inverted-U relationship between rewards and dishonesty that resembles our experimental results so long as the costs of lying are sufficiently convex (e.g. $b$ = 2.4). Although these numerical proofs and simulations do not cleanly establish the mechanism driving our results, they strongly suggest that a reference-dependent utility with convex costs of dishonesty magnitude are a possible explanation.

## 3. Study 2

To further explore whether prospect theory is a likely explanation for our results, we conducted a follow-up survey (see Appendix for details) with the 320 participants from Study 1. We wished to understand whether a likely reference point (daily MTurk earnings) or explicit stopping criterion was consistent with a prospect theory explanation for our data. As we noted earlier, these reference points are somewhat arbitrary and less desirable than more objective values (such as in Allen et al., 2016), but they are consistent with reference points used in prior field studies of effort allocation (Camerer et al., 1997; Fehr and Goette, 2007). The follow-up survey provides us an estimate of an Indian worker's average daily income on MTurk, their daily target income, and a stopping income level (if any) at which a worker is likely to quit their work for that day. 320 workers were invited to participate in the survey for a fixed fee of \$0.50; 239 workers completed the survey. We conducted Study 2 approximately three months after Study 1, a gap large enough to present Study 2 as an independent survey.

Table 2 provides descriptive statistics from participants in the follow-up survey. The male to female ratio remained similar to the bonus task, with a small drop of 1.9% from Study 1. The mean headcount for the sample in Study 2 is nearly the same at 3.93 compared to 3.91 for the original sample of 320 workers. To address outliers in reported income, target, and stopping point that likely represented errors (see Appendix), we winsorized worker responses on incomes at 5th and 95th percentiles. This ensured that outlier responses were not dropped, and instead assumed values from the data itself, at

**Table 2**
Descriptive Statistics for Study 2 Participants.

| Summary statistics | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| Gender (male = 1) | 239 | 0.67 | 0.47 | 0.00 | 1.00 |
| Age | 239 | 33.14 | 9.31 | 20.00 | 71.00 |
| Headcount | 239 | 3.91 | 1.35 | 0.00 | 6.00 |
| Total income | 239 | 12.39 | 6.89 | 1.50 | 31.50 |
| Before winsorizing | | | | | |
| Avg daily income (\$) | 239 | 6.68 | 16.04 | 0.08 | 230.77 |
| Target income (\$) | 239 | 9.34 | 12.85 | 0.01 | 153.85 |
| Stopping income (\$) | 239 | 18.79 | 69.37 | 0.00 | 1000.00 |
| After winsorizing | | | | | |
| Avg daily income (\$) | 239 | 5.41 | 5.33 | 0.50 | 20.00 |
| Target income (\$) | 239 | 8.50 | 7.60 | 1.00 | 30.00 |
| Stopping income (\$) | 239 | 12.46 | 13.47 | 0.15 | 50.00 |

*Notes:* Summary statistics for workers who participated in the follow-up survey on daily incomes. In order to address outliers in self-reported income, we winsorized average daily income, target income and stopping income at the 5th and 95th percentiles. Winsorized values helped ensure that outlier responses were not dropped, instead these outlier responses assumed values from the data itself. See study 2 in Appendix for details of survey on daily incomes.

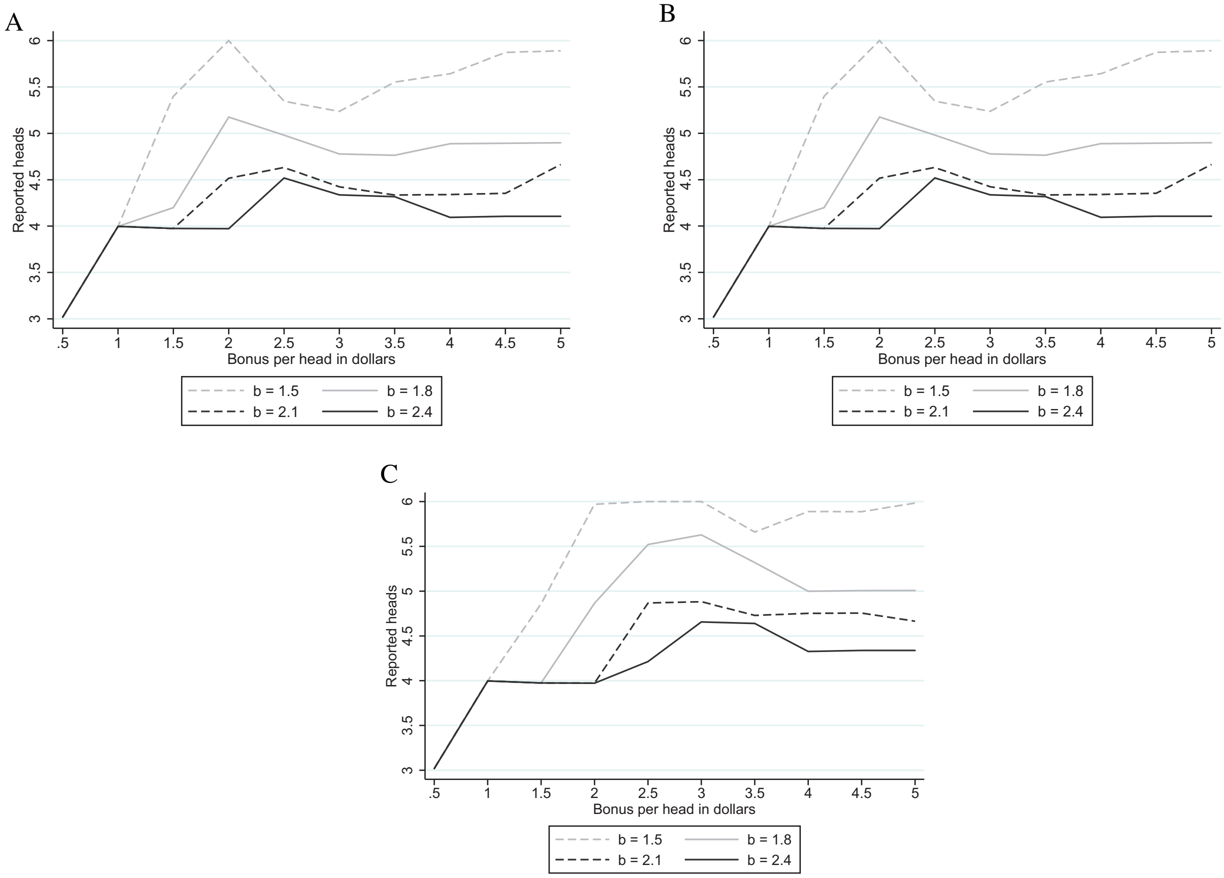


**Fig. 11.** (A) Simulated Predictions with Reference-Dependent Utility, Convex Costs of Lying Magnitude, and Reference Point of $8.
*Notes:* This figure shows results of the simulation of reference-dependent utility function with convex costs of lying $b > 1$, and for reference income point of $8 per day. The utility function was run on STATA using 10,000 simulated participants with randomly assigned coin flips and reward conditions. For sufficiently high convex costs of lying (e.g. $b = 2.1$ or $b = 2.4$), we observe a non-monotonic inverted-U relationship between rewards and dishonesty, similar to the experimental results from study 1.
(B) Simulated Predictions with Reference-Dependent Utility, Convex Costs of Lying Magnitude, and Reference Point of $12.
*Notes:* This figure shows results of the simulation of reference-dependent utility function with convex costs of lying $b > 1$, and for reference income point of $12 per day. The utility function was run on STATA using 10,000 simulated participants with randomly assigned coin flips and reward conditions. For sufficiently high convex costs of lying (e.g. $b = 2.1$ or $b = 2.4$), we observe a non-monotonic inverted-U relationship between rewards and dishonesty, similar to the experimental results from study 1.
(C) Simulated Predictions with Reference-Dependent Utility, Convex Costs of Lying Magnitude, and Reference Point of $16.
*Notes:* This figure shows result of the simulation of reference-dependent utility function with convex costs of lying $b > 1$, and for reference income point $16 of per day. The utility function was run on STATA using 10,000 simulated participants with randomly assigned coin flips and reward conditions. For sufficiently high convex costs of lying (e.g. $b = 2.4$), we observe a non-monotonic inverted-U relationship between rewards and dishonesty, similar to the experimental results from study 1.

the 5th and 95th percentiles.[7] After winsorizing for each worker, average daily income on MTurk is $5.41, mean daily target income is $8.50, and mean stopping income is reported as $12.46.

Fig. 12 shows the relationship between dishonesty and reward condition based on self-reported stopping income. Worker responses were split based on median self-reported stopping income, with stopping income from $0.15–$7 per day in the low group and $7.69–$50 in the high group. Black lines represent lowess fits and light grey lines are the raw data. We observe that within the sample of Study 2 participants, the non-monotonic inverted-U relationship is appears to be driven by those with lower stopping income, although our small sample limits precise identification of non-monotonicity (utest, p = 0.199). The figure illustrates that if one believes the stopping income level is an appropriate reference point, the reduced dishonesty at higher reward levels is likely driven by those who have already exceed their stopping point, which is consistent with our reference-dependent utility model in Eq. (4). Dishonesty falls in incomes that exceed the self-reported stopping income. The Appendix presents similar figures for daily income and target income.

[7] Winsorized values – average daily income ($0.5, $20), daily target income ($1, $30), and daily stopping income ($0.15, $50). Winsorization was necessary due to extremely outliers (e.g., $1000).

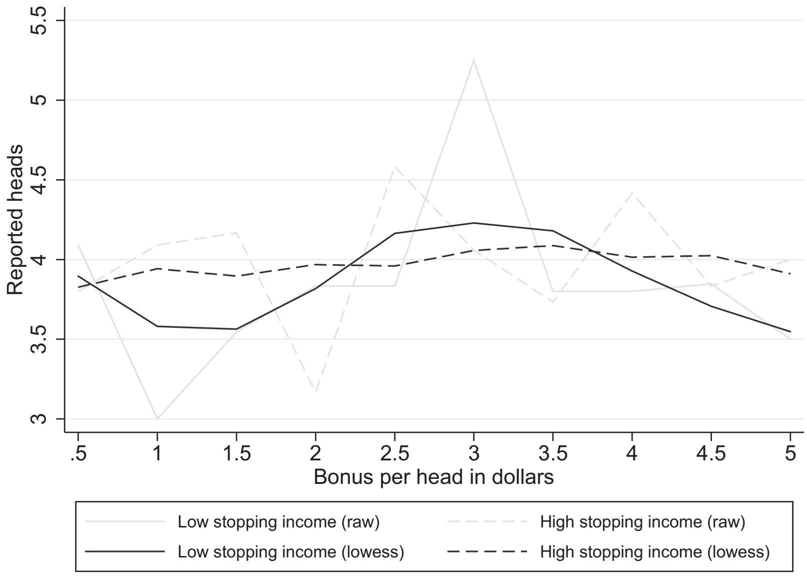


**Fig. 12.** Mean Reported Head Counts for Study 2 Participants with High and Low Reported Stopping Income.
*Notes:* This figure shows the relationship between reported heads and condition, grouped by level of self-reported stopping income in Study 2. Worker responses on stopping income were split using median responses among 2 groups. The black lines represent the lowess fit for both the low and high stopping income groups, while the light grey lines show the raw data. Within this sample, we observe that the lower stopping income group drives the non-monotonic inverted-U relationship between reported heads and condition.

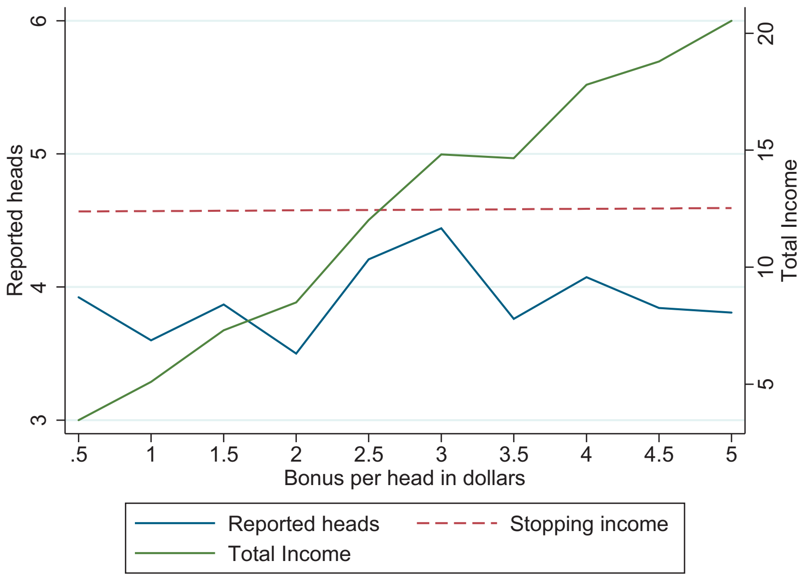


**Fig. 13.** Dishonesty Decreases After Surpassing Self-Reported Stopping Income.
*Notes:* This figure shows the mean reported heads and mean total income from Study 1, and overlaid with mean self-reported stopping income from Study 2. Total income from Study 1 includes $1.5 endowment for the OCR text recognition task and rewards from the bonus task. The dashed line represents a linear fit between self-reported stopping income and reward conditions. Workers reported mean stopping income of $12.46 per day (after winsorizing), as shown in Table 3. We observe reduced dishonesty levels in the higher incentive conditions, once total income surpasses mean stopping income in the figure above. Refer to question 4 in Study 2 for details. Responses in INR were standardized to USD using a conversion factor of 65.

Similarly, plotting reported head counts, total income (including $1.50 for the image recognition task), and average stopping income together in Fig. 13 shows a similar effect. The higher reward conditions that surpass average stopping income are those where dishonesty decreases. Although this is also circumstantial evidence for a reference-dependent utility, it is consistent with such an explanation. Furthermore, we note that we cannot rule out other unobservable individual characteristics such as guilt propensity that might be correlated with both stopping income and high-value theft.

## 4. Study 3

Although our use of a third-party website for the coin flip task makes individual dishonestly impossible to detect, we cannot fully rule out that participants erroneously perceive their behavior to be observable. If they indeed worried about their dishonesty being detected, and consequently not being paid, then one could explain our results in the same way as Kajackaite and Gneezy (2017)—the highest rewards reduce dishonesty through fear of detection.

We address this concern by surveying 644 Indian MTurk workers who were randomized into two sets of coin flip tasks. In the first set, participants were asked to flip a coin six times at random.org (similar to the bonus task in study 1) while in the second set, participants were asked to physically flip a real coin six times at home. After the coin flip task in either set, workers were presented questions in the form of a scenario in which other hypothetical participants received one of ten incentive levels ($0.50–$5.00 per head). Similar to Study 1, we randomly assigned participants to one of these ten hypothetical incentive-level conditions (in either set), while paying all participants the same fixed fee. Given the incentive level, participants were asked to provide a percentage estimate of hypothetical workers who would dishonestly over-report the number of heads in the coin flip task. They were also asked the average number of heads that a dishonest worker was likely to report (see Appendix for details).

A non-parametric Wilcoxon-Mann-Whitney test shows that the 324 workers using random.org and the 320 who flipped at home had statistically indistinguishable distributions of head counts ($p = 0.58$). Mean reported heads were 3.36 in the random.org task and 3.43 in the flip at home task—both lower than the 3.93 average in Study 1. We note that some participants still appear to be lying despite have no financial incentive to do so—the 3.39 average is still identifiably higher than the honest expectation of 3. We find no relationship, however, between the hypothetical incentive level and expectations about lying. Participants do not anticipate more frequent dishonest reports under higher incentives ($\beta_{condition} = 0.567$, $p = 0.510$), nor do they expect larger lies ($\beta_{condition} = 0.034$, $p = 0.225$). Expected dishonesty is largely consistent across conditions, with slightly less dishonesty expected in the $4 condition, probably because the randomly-generated true head counts of the participants were highest in this condition.

Most importantly, Study 3 provides insight about the expectations of being paid when reporting either honestly or dishonestly under each incentive level. Workers seem to have a higher certainty of payment when reporting truthfully (mean = 4.09/5), while there is moderate concern over payments when reporting dishonestly (mean = 3.60/5).[8] Although this suggests some concern about lie detection, we find no evidence that this is linked to incentive levels. In the random.org design, we find no statistically significant relationship between certainty scores and reward condition ($\beta_{certainty} = 0.056$; $p = 0.207$), or between self-reported concern scores and reward condition ($\beta_{concern} = -0.0241$; $p = 0.58$). For participants in the flip at home condition, we do not find a statistically significant relationship between certainty scores and reward condition ($\beta_{certainty} = -0.017$; $p = 0.693$), but a weak effect for the relationship between concern score and reward condition ($\beta_{concern} = -0.0865$; $p = 0.065$). However, we do find a statistically significant negative relationship between certainty scores and number of heads reported by participants ($\beta_{certainty} = -0.078$; $p = 0.002$). But the negative relationship for certainty of payment and number of heads reported is similar for both random.org ($\beta_{certainty} = -0.0809$; $p = 0.028$) as well as when flipping at home ($\beta_{certainty} = -0.0766$; $p = 0.036$). Results based on self-reported scores suggest that while workers may have some concern when reporting higher number of heads in either design conditions, this is uncorrelated with condition. These results suggest our Study 1 results cannot be explained by fear of detection and punishment. Descriptive statistics and additional results from Study 3 are presented in the appendix.

## 5. Discussion

In this paper we have provided evidence that the monotonicity of the relationship between dishonesty and rewards depends on the range of the financial stakes. At low incentive levels, incentives may have the null effect found in prior work (Mazar et al., 2008), but higher reward levels suggest a more complex relationship. Dishonesty peaks in conditions with substantial gains from dishonesty ($2.50–$3/head), then decreases as incentives grow to $5. Although we cannot directly observe the mechanism explaining this inverted-u relationship, our results are consistent with a prospect-theory based model where income above daily reference points motivates workers to eschew dishonesty due reduced benefits from rewards and convex costs of dishonesty. Our results are inconsistent with standard economic models of utility that are concave in long-term income, regardless of the functional form of the cost of dishonesty. Furthermore, our results support arguments that many individuals choose to cheat less than the maximum possible amount, even when the threat of detection is negligible. In our data, the cost of dishonesty is clearly internal, and dependent on the magnitude of the lie.

We emphasize that our reference-dependent utility explanation is largely based on our experimental design isolating participants from several other important factors that might produce a non-monotonic relationship between rewards and dishonesty. Kajackaite and Gneezy (2017) argue that their non-monotonic relationship results from fear of detection. A recent field experiment by Dugar and Bhattacharya (2017) explains decreased cheating in a fish market at higher prices by

[8] Based on self-reported certainty scores from 1-Very uncertain to 5-Very certain, and self-report concern scores from 1-Not at all concerned to 5-Extremely concerned

other-facing preferences. Our paper can neither refute nor confirm those explanations, which undoubtedly play important roles in shaping the reward-dishonesty relationship in more complex field and laboratory settings.

We note that our sample of Mechanical Turk workers, while not equivalent to a firm-based work force, is likely as generalizable to firm settings as university-based participant pools. MTurk represents a part-time or full-time job for most participants, and the financial stakes in our study are not trivial for their livelihood. In this sense, while we are cautious to generalize our results to the workplace, we believe they present a model of worker behavior that is at least as believable as many laboratory studies. Given Amir and Rand's (2012) conclusion that both laboratory and MTurk populations produce similar incentive responses at low levels, our results likely generalize to future lab studies as well. While it is possible that cultural differences can influence our results, it is important to remember that India itself is rapidly growing and a large economy, and shares cultural traits with its South Asian neighbors. With a population of over 1.25 billion, our study at the lower bound makes a contribution to a sizeable number of the global population. In this sense it is certainly no less generalizable that laboratory studies using European or American populations.

We acknowledge that it is nearly impossible for us to know workers' private beliefs about the risk of being caught, so we cannot fully dismiss that some participants erroneously believed they were being monitored. However, several elements of our research design make an explanation of detectability fears increasing with incentives unlikely. First, we used a third-party website that made our observation of dishonesty impossible. Second, even the highest payoffs from reporting six heads represent believable outcomes. Third, the counterfactual binomial distribution allows workers with low outcomes to cheat a little while actually *increasing* the likelihood of the outcome, thus decreasing fears of detectability. Finally, our third study, although purely hypothetical, does not suggest that fear of detection could explain our results across conditions. Indeed, respondents indicate some fear of detection, but that fear is not correlated with reward magnitude.

We also caution readers not to extrapolate the relationship found in our range of rewards to much higher levels not observed here. Suri et al. (2011), using an MTurk sample from both India and the U.S., found limited evidence of cheating and no relationship with reward level. Their payoffs, however, were 5–10 times smaller than ours. In the same sense that our highest levels present different conclusions that previous work with lower incentives, so too might future work that extends rewards by an order of magnitude. But we believe our results emphasize that commonly held beliefs that individuals cheat "a little but not a lot" oversimplify the importance of reward structure in dishonesty. Clearly, this question begs for more research in field settings and in higher stakes experiments.

Finally, given that we cannot identify the mechanisms driving our non-monotonic relationship, we hope our results will motivate future work to precisely identify the relationship between incentives and both the value of income as well as the internal cost of dishonesty. We have presented one possible explanation—reference-dependent utility with convex internal costs—but our evidence of this is purely speculative and circumstantial. Future work should attempt to directly manipulate reference points or expectations to identify whether reference-dependent utility can truly reduce dishonesty at higher levels.

## Acknowledgements

The Center for Research in Economics & Strategy funded this study. We are grateful to Srinidhi Balasubramanian, Hengchen Dai, Joe Goodman, Rick Larrick, Nina Mazar, and Jason Snyder for help in developing this paper.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jebo.2017.03.022.

## References

Abeler, B.J., Falk, A., Goette, L., Huffman, D., 2011. Reference points and effort provision. Am. Econ. Rev. 101 (2), 470–492.
Abeler, J., Becker, A., Falk, A., 2014. Representative evidence on lying costs. J. Public Econ. 113, 96–104.
Allen, E.J., Dechow, P.M., Pope, D.G., Wu, G., 2016. Reference-dependent preferences: evidence from marathon runners. Manage. Sci. (forthcoming).
Amir, O., Rand, D.G., 2012. Economic games on the internet: the effect of $1 stakes. PLoS One 7 (2), e31461.
Andersen, S., Ertaç, S., Gneezy, U., Hoffman, M., List, J.A., 2011. Stakes matter in ultimatum games. Am. Econ Rev. 101 (7), 3427–3439.
Barberis, N.C., 2013. Thirty years of prospect theory in economics: a review and assessment. J. Econ. Perspect. 27, 173–195.
Becker, G.S., 1968. Crime and punishment: an economic approach. J. Polit. Econ. 76, 169–217.
Bennett, V., Pierce, L., Snyder, J. a., Toffel, M.W., 2013. Customer-driven misconduct: how competition corrupts business practices. Manage. Sci. 59, 1725–1742.
Branco, F., Villas-Boas, J.M., 2015. Competitive vices. J. Mark. Res. 52, 801–816.
Camerer, C.F., Hogarth, R.M., 1999. The effects of financial incentives in experiments: a review and capital-labor-production framework. J. Risk Uncertain. 19, 7–42.
Camerer, C., Babcock, L., Loewenstein, G., Thaler, R., 1997. Labor supply of New York City cabdrivers: one day at a time. Q. J. Econ. 112, 407–441.
Carpenter, J., Verhoogen, E., Burks, S., 2005. The effect of stakes in distribution experiments. Econ. Lett. 86, 393–398.
Cherry, T., Frykblom, P., Shogren, J., 2002. Hardnose the dictator. Am. Econ. Rev. 92 (4), 1218–1221.
Cohn, A., Fehr, E., Maréchal, M., 2014. Business culture and dishonesty in the banking industry. Nature 516, 86–89.
Cuzick, J., 1985. A wilcoxon-type test for trend. Stat. Med. 4, 543–547.
Dreber, A., Johannesson, M., 2008. Gender differences in deception. Econ. Lett. 99, 197–199.
Dugar, S., Bhattacharya, H., 2017. Fishy behavior: a field experiment on (dis) honesty in the marketplace. J. Behav. Exp. Econ. 67, 41–55.
Engström, P., Nordblom, K., Ohlsson, H., 2015. Tax compliance and loss aversion. Am. Econ. J. Econ. Policy 7, 132–164.
Erat, S., Gneezy, U., 2012. White lies. Manage. Sci. 58, 723–733.

Fehr, E., Goette, L., 2007. Do workers work more if wages are high? Evidence from a randomized field experiment. Am. Econ. Rev. 97 (1), 298–317.
Fischbacher, U., Föllmi-Heusi, F., 2013. Lies in disguise – an experimental study on cheating. J. Eur. Econ. Assoc. 11, 525–547.
Gächter, S., Schulz, J.F., 2016. Intrinsic honesty and the prevalence of rule violations across societies. Nature 531, 496–499.
Garbarino, E., Slonim, R., Villeval, M.C., 2016. Loss aversion and lying behavior: theory, estimation and empirical evidence. Working paper, Available at SSRN: https://ssrn.com/abstract=2875989 or http://dx.doi.org/10.2139/ssrn.2875989.
Gino, F., Pierce, L., 2009. Dishonesty in the name of equity. Psychol. Sci. 20, 1153–1160.
Gino, F., Pierce, L., 2010. Lying to level the playing field: why people may dishonestly help or hurt others to create equity. J. Bus. Ethics 95, 89–103.
Gneezy, U., 2005. Deception: the role of consequences. Am. Econ. Rev. 95 (1), 384–394.
Goodman, J.K., Cryder, C.E., Cheema, A., 2013. Data collection in a flat world: the strengths and weaknesses of Mechanical Turk samples. J. Behav. Decis. Mak. 26, 213–224.
Goodman, J.K., 2015. How and when to Ask the Turk: Tips for Using Amazon's Mechanical Turk to Conduct Consumer Research. Workshop at the Meetings of the Association for Consumer Research. LA, New Orleans, October.
Grolleau, G., Kocher, M.G., Sutan, A., 2016. Cheating and loss aversion: do people cheat more to avoid a loss? Manage. Sci. 62, 3428–3438.
Heath, C., Larrick, R., Wu, G., 1999. Goals as reference points. Cogn. Psychol. 38, 79–109.
Ipeirotis, P., 2010. Demographics of Mechanical Turk. Work. Pap. CeDER-10-01. Available at SSRN: 10.2139/ssrn.1585030.
John, L.K., Loewenstein, G., Rick, S.I., 2014. Cheating more for less: upward social comparisons motivate the poorly compensated to cheat. Organ. Behav. Hum. Decis. Process. 123, 101–109.
Köszegi, B., Rabin, M., 2006. A model of reference-dependent preferences. Q. J. Econ. 121, 1133–1165.
Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. Econom. J. Econom. Soc., 263–291.
Kajackaite, A., Gneezy, U., 2017. Incentives and cheating. Games Econ. Behav. 102, 433–444.
Kartik, N., 2009. Strategic communication with lying costs. Rev. Econ. Stud. 76, 1359–1395.
Kern, M.C., Chugh, D., 2009. Bounded ethicality: the perils of loss framing. Psychol. Sci. 20, 378–384.
Kilduff, G., Galinsky, A., Gallo, E., 2016. Whatever it takes to win: rivalry increases unethical behavior. Acad. Manag. J. 59, 1508–1534.
Lind, J.T., Mehlum, H., 2010. With or without u? The appropriate test for a u-shaped relationship. Oxf. Bull. Econ. Stat. 72, 109–118.
Litman, L., Robinson, J., Rosenzweig, C., 2015. The relationship between motivation, monetary compensation, and data quality among US- and India-based workers on Mechanical Turk. Behav. Res. Methods 47, 519–528.
Litman, L., Robinson, J., Abberbock, T., 2016. TurkPrime.com: a versatile crowdsourcing data acquisition platform for the behavioral sciences. Behav. Res. Methods, 1–10.
Lowes, S., Nunn, N., Robinson J., Weigel, J., 2017. The evolution of culture and institutions: Evidence from the Kuba Kingdom Econometrica, (forthcoming).
Mazar, N., Aggarwal, P., 2011. Greasing the palm: can collectivism promote bribery? Psychol. Sci. 22, 843–848.
Mazar, N., Amir, O., Ariely, D., 2008. The dishonesty of honest people: a theory of self-concept maintenance. J. Mark. Res. 45, 633–644.
Peer, E., Acquisti, A., Shalvi, S., 2014. "I cheated, but only a little": partial confessions to unethical behavior. Journal of Personality and Social Psychology 106, 202–217.
Rabin, M., 2000. Risk aversion and expected-utility theory: a calibration theorem. Econometrica 68, 1281–1292.
Rees-Jones, A., 2017. Quantifying loss-averse tax manipulation. Available at SSRN 2330980.
Schweitzer, M.E., Ordóñez, L., Douma, B., Ordonez, L., 2004. Goal setting as a motivator of unethical behavior. Acad. Manag. J. 47, 422–432.
Shalvi, S., Dana, J., Handgraaf, M.J.J., De Dreu, C.K.W., 2011. Justified ethicality: observing desired counterfactuals modifies ethical perceptions and behavior. Organ. Behav. Hum. Decis. Process. 115, 181–190.
Shalvi, S., Eldar, O., Bereby-Meyer, Y., 2012. Honesty requires time (and lack of justifications). Psychol. Sci. 23, 1264–1270.
Shu, L.L., Gino, F., Bazerman, M.H., 2011. Dishonest deed, clear conscience: when cheating leads to moral disengagement and motivated forgetting. Pers. Soc. Psychol. Bull. 37, 330–349.
Suri, S., Goldstein, D., Mason, W., 2011. Honesty in an online labor market. Hum. Comput. 11, 61–66.
Sutter, M., 2009. Deception through telling the truth?! Experimental evidence from individuals and teams. Econ. J. 119, 47–60.
Tversky, A., Kahneman, D., 1992. Advances in prospect theory: cumulative representation of uncertainty. J. Risk Uncertain. 5, 297–323.
Wang, L., Murnighan, J.K., 2016. How much does honesty cost? Small bonuses can motivate ethical behavior. Manage. Sci. (forthcoming).
Weisel, O., Shalvi, S., 2015. The collaborative roots of corruption. Proc. Natl. Acad. Sci. 112, 10651–10656.