HONESTY ON THE MARGINS

SIMON D. HALLIDAY, JONATHAN LAFKY, AND ALISTAIR J. WILSON

Abstract. We examine how competition affects honesty through two channels: an intensive margin (endogenizing the rewards to exaggeration) and an extensive margin (endogenizing the pool of competitors via an outside option). Using the lying and reputation-cost preference model from the Abeler et al. (2019) metastudy we predict minimal effects on honesty from either channel in isolation, but a sharp decrease in honesty with both competitive channels. Our experimental results confirm the predictions of the model, over both reports and entry/exit behavior. However we do detect some anomalies, suggesting that smaller communities/groups can have persistent honesty even with both competitive forces.

1. Introduction

Selection of the most capable into particular professions via healthy competition is a mainstay of economic analysis, often resulting in desirable features: encouraging and allocating effort, rewarding innovation, and increasing efficiency. But where market forces and morality collide, competition and selection can be actively harmful. An example of this is honesty. If competition increases the temptation to stretch the truth, or if the market selects for those with the fewest moral objections, then competition can result in worse market outcomes. For example, a job seeker in isolation might gently embellish their résumé, but feel compelled to engage in wholesale fictions if they believe other candidates are doing the same. Salespeople and marketers can face an incentive to exaggerate their product's qualities, causing those uncomfortable with commission-enhancing fabrications to select into other professions. The end effect of these competitive margins can therefore be an efficiency-decreasing collapse of faith in cheap talk statements.

By contrast, a prominent behavioral literature has shown that many decision makers have strong moral imperatives, willing to give up pecuniary rewards to be (and be perceived as) honest. This moral aversion to lying can produce relatively honest interactions even in settings where exaggeration yields tangible rewards. However, lying averse preferences

Date: February, 2025.

Our thanks to David Huffman, Collin Raymond and Lise Vesterlund for comments. Halliday: Center for Economy and Society, Johns Hopkins University; simon.halliday@jhu.edu. Lafky: Department of Economics, Carleton College; jlafky@carleton.edu. Wilson: Department of Economics, University of Pittsburgh; alistair@pitt.edu.

are often examined in isolation, but as motivated above, the relative rewards for dishonesty are often endogenous (increasing in the dishonesty of others), and selection forces can endogenize the pool (selecting out those with the most-intense moral qualms). How do these two market forces affect honesty under the models of lying aversion identified in the behavioral literature?

In this paper, we theoretically and experimentally examine these two competitive margins on honesty. We start out with a baseline setting with fixed rewards to honesty and no selection—using the well-studied Fischbacher and Föllmi-Heusi (2013) die-roll reporting paradigm. We then compare this baseline to environments where we add: (i) competition across the reports, (ii) selection out of the competition, and (iii) both market forces together. Our baseline environment assigns fixed prizes for each report using lotteries, using a linearly increasing chance of winning a prize, allowing us to interpret the environment as one where the agent competes for the highest report against an honest opponent. Our competition treatments then examine the effects of endogenizing the payoffs from each report, matching participants with one another in a report tournament and assigning a prize to the highest report. As such, the tangible benefits of a lie (as well as any psychological costs) will depend on the reports of others. Our selection treatments modify the environment by providing an outside-option task, with a fixed probability of winning the prize that does not depend upon the participant's honesty. As such, those with moral qualms about lying can opt out of the task, but in so doing alter the pool of those that remain. Finally, in our treatments with both market forces, we allow for both competition (endogenizing the expected prize for each report) and selection (endogenizing the pool of those competing).

The idea that honesty suffers in competitive environments is intuitive; however, the mechanisms driving this can be nuanced. A growing experimental literature has served to identify the features of preferences driving honest behavior via isolated decision problems and clever identification. Lying-averse preferences have been extensively documented—the excellent Abeler et al. (2019) meta study nicely distills the literature's findings. While many individuals show an *innate* aversion to acting dishonestly, there is also: (i) a great deal of heterogeneity across individuals; and (ii) substantial concern for reputation, the desire to avoid the appearance of being a liar. A natural question is whether this type of preference can help us understand the effects of competition and selection.

Embedding the meta-study calibrated "lying-cost plus reputation" model into our environment, we examine the effects of each market margin in turn, and then together. The preference model posits two behavioral components within a population: a fixed psychological lying cost from making any dishonest report, and a reputation cost for each report that endogenously depends on an outside observer's equilibrium perceptions that each report is a lie, with varying intensity across individuals. Retaining the population-level equilibrium approach, we derive predictions when we endogenize the prize structure through our report tournament (making the prize structure more convex), endogenizing the population (where individuals with reputation costs above an equilibrium cutoff opt out of the market), and then both forces together.

Using the meta-study calibrated model, we can make exact quantitative predictions, however, the model also provides for intuition to understand the effects. Endogenizing the prize structure has three main effects: (i) The linear payoff in the baseline becomes more convex, where the marginal returns to higher reports become relatively larger. (ii) Reputation costs provide a moderating force, inhibiting the number of maximal reports from growing too large. (iii) Expected utility is reduced, thanks to a reduction in the chance of winning as competitors become dishonest, as well as an increase in reputational costs from lying.

When we endogenize the population with a fixed outside option, the predictions are relatively straightforward. In equilibrium, the expected utility will be decreasing in the intensity of the reputation cost. As such, if a fixed outside option is available, selection will be for those with the highest reputation costs to opt out. For low outside options this will not have much of an effect, but as the value of the outside option increases the effects can be more substantive, weakening the reputation forces that stop participants from making a maximal report. While selection does increase the reputation costs of those that remain in the market, the fixed prizes for each report (assuming competitors are completely honest) stop the expected utility from being decreased too much.

However, the combination of both market forces leads to substantial dishonesty. Full competition over the reports reduces expected utility through both decreased prizes and increased costs in the market. For any fixed outside option, the decreases in expected utility within the market leads to even greater adverse selection within the market, where the moderating forces from reputation are greatly diminished as the participants that remain are those who feel the lowest reputation costs. Even where outside options are terrible, the effects on honesty from both market forces at once can be substantial, while more moderate outside options can lead to complete dishonesty.

Our experimental design examines the large-population predictions from the Abeler et al. (2019) 'LC+reputation' preference model across six experimental treatments in a 3×2

design. Treatments vary across the intensity of the outside option (not available, a low payoff, and a high payoff) and the presence of competition over reports (with participant reports compared pairwise against either an honest robot, or another participant). Using calibrated preferences from the Abeler et al. (2019) meta-study, we examine the predictions in each treatment with three broad predictions: on its own, neither (i) selection nor (ii) competition substantially alter the aggregate predicted honesty relative to the baseline (no competition, no outside option). However, (iii) when both forces are combined, the prediction is for a substantial reduction in honesty.

Our experimental results find strong support for the equilibrium predictions. Neither pure selection nor pure competition cause substantial changes in reporting behavior. However, the interaction of both forces leads to a very substantial increases in dishonesty. Although the aggregate effects are large, we also see considerable heterogeneity among our experimental groups. While many of the experimental groups exposed to both market forces devolve to nearly complete dishonesty, some groups do manage to converge to a stable outcome with substantial honesty levels. Contrary to the theoretical prediction, we actually see greater honesty in the setting with the higher outside options. The reason for this anomaly is that the large population assumption we use to understand the environment lacks bite in our setting with high outside option, as the rematching pool becomes smaller. While outside the scope of the model, the results here indicate one potential silver lining: in small enough communities honesty might persist even with both market forces.

In terms of organization, our paper next outlines the surrounding literature. In Section 2 we discuss the specifics of our experimental design, the treatments, and what the LC+Reputation model predicts in each. Section 3 outlines the experimental results and Section 4 concludes.

1.1. **Literature.** An extensive literature has shown the decision to lie depends not only on material payoffs, but also other factors such as the "size of the lie" and image-based concerns for appearing to be dishonest (Gneezy, 2005; Mazar et al., 2008; Gneezy et al., 2013, 2018). Past studies have found mixed effects of competition on dishonesty, often in the context of reporting output in real-effort experiments. Belot and Schröder (2013) and Cadsby et al. (2010), find more overreporting of output under competition than piece-rate compensation, while Carpenter et al. (2010) show that competitive incentives cause subjects to dishonestly underreport their coworkers' productivity. Similarly, Schwieren and Weichselbaumer (2010) find that competition results in a modest increase in dishonesty in a maze-solving task, driven largely by increases in cheating among low-performing subjects, while Rigdon and D'Esterre (2015) find competition has no effect when reporting one's own productivity, and actually *increases* honesty when reporting the output of competitors. In tournament-style competitions, Necker and Paetzel (2023) find no increased dishonesty relative to random matching, Conrads et al. (2014) show that dishonesty is responsive to the size of prize, with dishonesty growing as the relative reward for winning increases, while Cartwright and Menezes (2014) show a non-monotonic relationship between dishonesty and the intensity of competition; dishonesty is highest with moderate levels of competition, as lying is unnecessary in a low-competition environments and useless in high-competition ones.

Several aspects of our environment have been previously studied, but, to our knowledge, never in conjunction with one another. Faravelli et al. (2015) examine selection and competition in a one-shot environment, finding that exogenously imposed competition increases dishonesty, but endogenous competition results in no net change in outcomes. Olsen et al. (2019), Hanna and Wang (2017) and Barfort et al. (2019) show an individual's preference for dishonesty can predict selection into public-sector jobs, where less-honest candidates can be either more or less interested, depending on the country studied. Benistant et al. (2022) and Feltovich (2019) compare repeated interactions with and without competition, but with no choice of endogenous entry, both finding increased dishonesty under competition. Konrad et al. (2021) and Fehrler et al. (2020) both show that dishonest subjects will pay a substantial premium to enter a task that rewards dishonesty. Focusing on the endogeneity of the rewards to lying, Dannenberg and Khachatryan (2020) find no difference in dishonesty when a die-rolling competition is against another subject than when it is against a guaranteed honest roll by the experimenter. Casella et al. (2018) find that competition between senders in a sender-receiver game results in more dishonest exaggerations by senders, but does not change receiver behavior or final outcomes.

Central to our contribution is our focus on the long-run effects of competition on dishonesty, allowing time for dynamic adjustment over both the choice to enter the competition and the decision of how honestly to compete. There is mixed evidence as to how repetition itself can influence the decision to behave honestly. Fischbacher and Föllmi-Heusi (2013) find that repetition leads an almost 50% increase in maximal or near-maximal reports, while d'Adda et al. (2017) find largely stable behavior far below maximal dishonesty, similar to what they observe in a one-shot environment. In their meta-study, Abeler et al. (2019) find generally modest effects from repetition, with subjects initially slightly more honest in repeated environments than in one-shot settings, though behavior trends toward the one-shot level over time.

2. Experimental Design

2.1. **Basic Environment.** Our experiment implements a two-by-three factorial design with variation over: (i) The nature of the matched outcome in the competitive task, either an exogenous response from the computer or an endogenous response from another participant. (ii) The value of the expected reward in the alternative outside-option task, either with no outside option, a low-value reward, or a high-value reward. Participants make choices across 30 identically structured periods in order to allow for convergence,[1] where in each period they must report on the privately observed outcome from a ten-sided die roll. Each period has the following sequence:

(i) Choice of Task: The participant selects their task: either the *fixed* task with a constant expected return for each die-roll report (but where the report is still instrumental in determining the realized outcome), or the *competitive* task with an expected return that increases with the die-roll report.

(ii) Report: After choosing their task, each participant privately rolls a 10-sided die with faces labeled zero to nine, and reports an outcome roll (an unconstrained choice of report, zero to nine).

(iii) Feedback: After all participants have reported an outcome, task payments are realized, and information is provided about the outcome.

While participants' reported rolls are used to determine outcomes in both tasks, the relationship between reports and expected payoffs differ by task. In the fixed task, participants are offered the lottery $\mathcal{L}_F(p) = p \cdot \$15 \oplus (1-p) \cdot \$5$, winning a \$15 prize with probability p and a \$5 prize with probability $1-p$. We implement this lottery experimentally via a draw from an urn with equal numbers of odd or even-labeled balls, where the outcome of the lottery depends on whether the reported roll matches the type of ball drawn from the urn. The participant's report combined with the drawn ball determine the outcome,

[1]Our choice to have a dynamic design, with multiple repetitions, is a function of our ambitions. We want to understand the equilibrium interactions between selection and competition. Without repetition, we did not see any plausible way in which equilibrium forces would be able to arise. Our underlying model requires agents to have a clear sense of the expected report distributions in the competitive task to make an informed choice of task. Expecting agents to be able to form a rational expectation over that distribution in a one-shot experiment seems unreasonable, where most interpretations of equilibrium theory will instead be for the steady state, after learning has been completed. Similar to the study of simple strategic games, while one-shot interactions can be important, studying convergent behavior is perhaps the only way that equilibrium behavior can be given a fair shake, where we note that our shift to a dynamic frame is across all of our treatments.

however the probability p of winning the \$15 prize does not vary with the participant's report.[2]

In contrast, participants' expected payoffs in the competitive task are increasing in the reported roll. Prizes for the competitive task are determined by comparing the reported roll x to a matched roll y, where the generating process for y (either exogenous or endogenous) varies by treatment. Competitive-task participants earn \$15 if their reported roll is the larger of the two, \$5 if their roll is smaller, and \$15 or \$5 with equal probability if the rolls are equal. A competitive-task report therefore produces a lottery $\mathcal{L}_C(x,y) = \pi(x,y) \cdot \$15 \oplus (1-\pi(x,y)) \cdot \$5$, where the probability of winning the \$15 prize given realized reports x and y is given by:

$$\pi(x,y) = \begin{cases} 1 & \text{if } x > y, \\ 1/2 & \text{if } x = y, \\ 0 & \text{if } x < y. \end{cases}$$

In the feedback stage, participants are informed of the outcome in the selected task. Fixed-task participants are reminded of their report x and shown the realized urn draw and the resulting prize. Competitive-task participants are similarly reminded of their report x, and are given feedback on the competing roll y and the resulting prize for the round.

2.2. **Treatments.** Our experiment was conducted over twelve sessions, each comprised of 24 unique undergraduate participants recruited at the Pittsburgh Experimental Economics Laboratory (PEEL).[3] Each session implemented a single treatment, resulting in a between-subject design where we vary the environment across two dimensions:

Nature of competition: Our first treatment variable manipulates the source of the y roll in the competitive task. In the *exogenous* treatments y is drawn from an independent discrete uniform distribution over $\{0,\dots 9\}$, equivalent to a fair roll of the

[2]In detail, after participants report their roll x in the fixed task, we draw a ball from a virtual urn with 100 balls: n balls labeled *Odd*, n labeled *Even*, 5 labeled *Any*, and $95-2n$ labeled *Neither*. The participant earns \$15 if the reported roll matches the drawn ball, and \$5 if it does not. As such, any reported roll $x \in \{0,2,4,6,8\}$ receives the \$15 prize if the drawn ball is *Even* or *Any*; while $x \in \{1,3,5,7,9\}$ receives the \$15 prize if the drawn ball is *Odd* or *Any*. The probability of winning the \$15 prize, one of our treatment variables, is therefore $p = (n+5)/100$ regardless of the participant's report, where the *Any* and *None* balls allow us to vary p without any other language changes.

[3]All sessions were conducted using the zTree software package (Fischbacher, 2007). Experimental instructions are available in the appendix. We maintain neutral language throughout the instructions, avoiding any reference to "dishonesty" or "lying," and referring to the two tasks simply as "Task A" and "Task B."

ten-sided die. Because the y roll is randomly determined, the exogenous treatments are individual decision problems. In contrast, our *endogenous* treatments use die-rolls from other participants as the source for y, and are more-formally games. In particular, we assign participants in all endogenous sessions to matching groups of size six, and draw the matched report y randomly from one of the other five group members who also entered the competitive task.[4]

Outside-option value: Our second treatment variable manipulates the expected value in the fixed task. The *L(ow)* treatments has a win probability of $p_L = 0.25$, while the *H(igh)* treatments have a win probability of $p_H = 0.45$. In our $\emptyset$ (*None*) treatment, the participants have no outside option, and the only possible task is the competitive one (so effectively $p_\emptyset = 0$).

We refer to our six treatments through the labels $\text{EX}_\emptyset$, EX_L, EX_H, $\text{EN}_\emptyset$, EN_L, and EN_H, as summarized in Table 1. The table includes the source of the y-roll, the value of the outside option (as a probability of winning the \$15 prize), followed by the number of sessions, matching groups, participants and decisions. In terms of recruitment, because participants did not interact with one another in the three exogenous treatments (and so form singleton matching groups) we conducted only one session of 24 participants for each. For our three endogenous treatments (with matching groups of size six) we conducted three separate sessions, and recruited 72 total participants for each endogenous treatment.

2.3. **Theory.** While we defer more formal hypotheses until after we outline the theory, the fundamental questions we seek to answer through our experiment are:

Question 1. *What are the effects of endogenizing the returns from competition?*

Question 2. *What are the effects of selection when we allow for opting out of the competition?*

Question 3. *Do these two economic forces interact in a substantial way?*

To make predictions and structure our analysis, we turn to the extensive experimental literature surrounding self-reports in the die-roll task. In particular, the meta-study by Abeler et al. (2019) consolidates this literature, examining which preference models best

[4]Matching is random, and not necessarily bilateral, where i being matched to j does not require that j is matched to i. If all others participants choose the fixed task, the draw y comes from a robot player playing honestly as in the EX treatments (in practice, 97.5 percent of competing draws were from other participants). All of this is common knowledge to participants.

Table 1. Experimental Design

(a) Treatment design

Treatment	Prize	Outside option	Sessions	Groups	Subjects	Decisions
$EX_\emptyset$	Fixed	None	1	24	24	720
EX_L	Fixed	0.25	1	24	24	720
EX_H	Fixed	0.45	1	24	24	720
$EN_\emptyset$	Competitive	None	3	18	72	2,160
EN_L	Competitive	0.25	3	18	72	2,160
EN_H	Competitive	0.45	3	18	72	2,160
Total			**12**	**126**	**288**	**8,640**

(b) Treatment predictions

Treatment	Entry	Avg. report	Honesty	Max Reports	Min Reports
$EX_\emptyset$	1	5.4	0.79	0.17	0.04
EX_L	1	5.4	0.79	0.17	0.04
EX_H	0.25	6.5	0.57	0.29	0.00
$EN_\emptyset$	1	5.4	0.80	0.19	0.05
EN_L	0.20	6.7	0.51	0.39	0.02
EN_H	ϵ	9.0	0.0	1.00	0.00

explain observed behavior. One conclusion from their analysis is that the main features of the meta-dataset can be explained via the following preference:

$$u_i(\omega, x; \theta_i) = \mathbb{E}\pi(x) - C \cdot \mathbf{1}_{\omega \neq x} - \theta_i \cdot \Lambda^\star(x).$$

The preferences model here comprises: $\mathbb{E}\pi(x)$, the expected prize associated with the report x; C, a fixed psychic cost to the individual from any report x that differs from the true state ω; and $\theta_i \cdot \Lambda^\star(x)$, a reputation cost incurred from making the report x (regardless of the veracity). The components of the reputation cost are the individual-level weight placed on reputation, θ_i, which is a private draw from the distribution $\mathcal{U}_\theta$ (a uniform draw on $[0, \kappa]$); and $\Lambda^\star(x)$, the fraction of participants that report x who are lying.[5]

The engine of the above preference model is the $\Lambda^\star(x)$ term, an equilibrium object that endogenizes the reputation cost for each report x as a function of the prize structure $\pi(\cdot)$ and the state space Ω. A report x with a large payoff $\pi(x)$ that is not associated with lying

[5]To make sure this environment is well defined in the limit, we supplement the above model by allowing for an arbitrarily small mass of payoff-maximizing types. That is, we allow for $(1-\epsilon)$ of the agents to have the Abeler et al. (2019) preferences, where $\epsilon > 0$ have a lying cost of $C = 0$ and a reputation cost of $\theta = 0$. These types will always report maximally in the competitive task, and make sure that limiting outcomes are well defined.

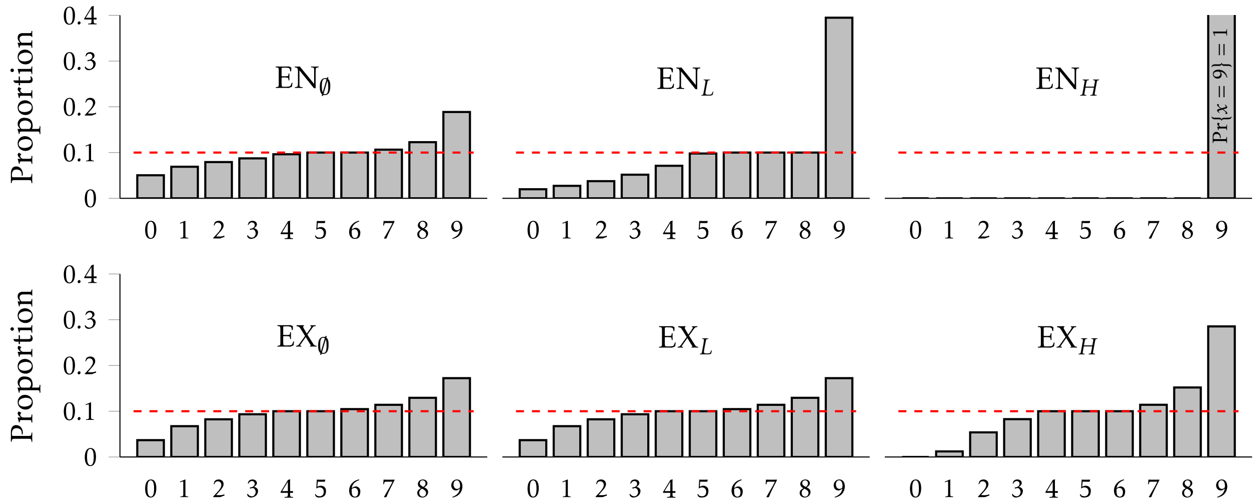

FIGURE 1. Theoretical report distributions

Note: Histograms show unconditional probabilities of each die roll 0 to 9. The dotted reference line shows the true data-generating process of a 1/10 probability.

will be more tempting to report, despite the psychic costs of lying, if it generates both a substantial payoff increase and is not associated with liars. However, such options will be impossible in an equilibrium, as the fraction of dishonest reports $\Lambda^\star(x)$ must reflect the expected tradeoffs between the gains from a lie at each report x with the reputation losses.

The above model (*lying costs plus reputation*, henceforth *LC+Rep*) is calibrated in Abeler et al. to the meta-study data. In our setting, their calibrated parameters translate to a lying cost of $C = 27/50$ and an upper bound on the reputation-cost support given by $\kappa = 2.16$.[6] Where possible, we use the general LC+Rep model to make qualitative arguments, directional predictions and provide intuition on the economic effects. However, given the substantial endogeneity introduced in our setting, we also make use of the specific calibration to make quantitative predictions to illustrate the scale of the effects.

In our baseline setting $\text{EX}_\emptyset$ (exogenous uniform draws for y, no outside option) the expected prize $\pi(x;y)$ is linearly increasing in the report x. Reporting $x = 0$ yields a 1/20 probability of winning, where each increment in the report increases the chance of winning by 1/10 (with the maximal 9-report yielding a 19/20 probability). Without an outside option, this is the standard die-roll paradigm. We can solve for the equilibrium reporting

[6]Their model is calibrated with $C = 3$ and $\kappa = 12$ over payoffs from 1 to 6. Our win probability in the base environment spans 1/20 to 19/20, so a payoff normalization leads to the given parameters.

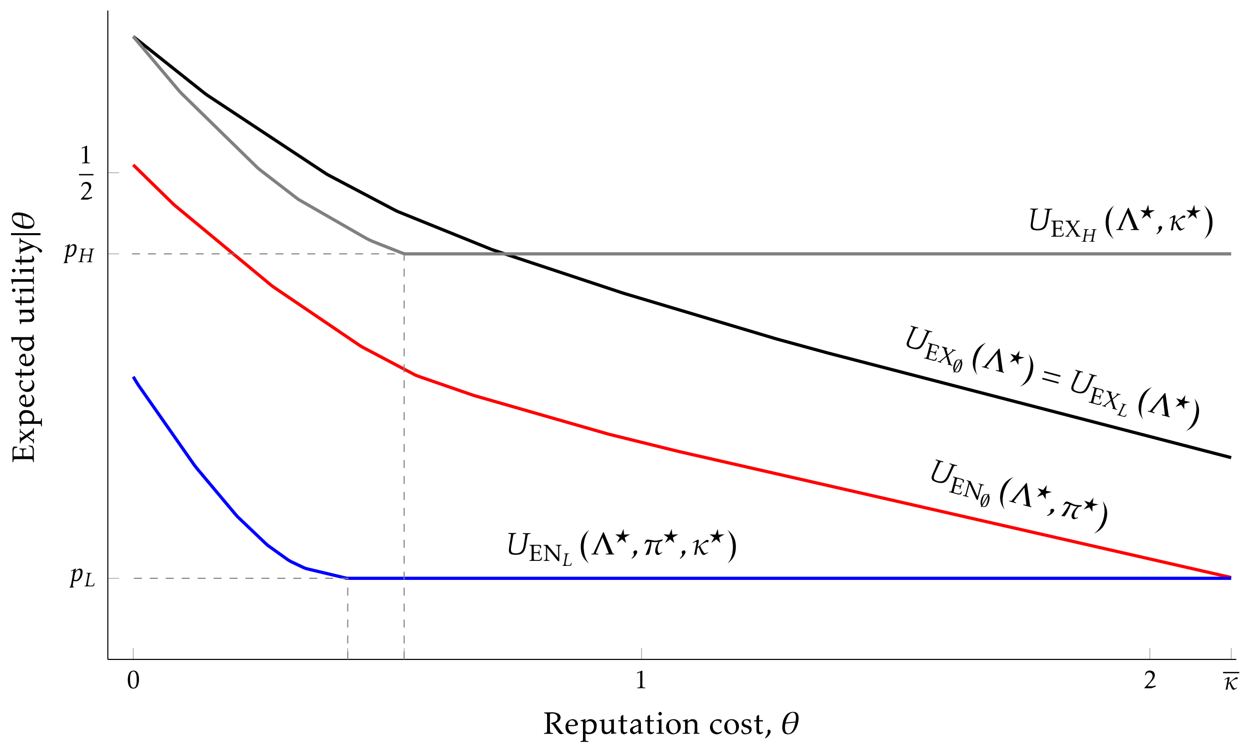

FIGURE 2. Expected utility by treatment

policy $\xi^\star : \Omega \times [0, \kappa] \to \Omega$, which maps the actual roll ω and the individual's reputation cost θ_i into a chosen report. The equilibrium report distribution for the $EX_\emptyset$ treatment using the literature-calibrated parameters is indicated in the bottom left graph of Figure 1—where for each report x we graph $\Pr\{\xi^\star(\omega,\theta) = x\}$. The figure illustrates the typical experimental pattern: substantial honesty, with some lying at both the payoff-maximizing report (9), but also at high but not-maximal reports (here 7 and 8).

Before examining the effects from competition and selection, we return to the model and outline how the expected utility in equilibrium varies across the different reputation costs. The expected utility across the reputation types in $\mathrm{EX}_\emptyset$ is given by:

$$U_{\mathrm{EX}_\emptyset}(\theta) = \mathbb{E}_\omega\left[u\left(\omega, \xi^\star(\omega,\theta); \Lambda^\star, \pi, \kappa\right)\middle|\theta\right].$$

So long as there is some dishonesty ($\exists x \in \Omega$ such that $\Lambda^\star(x) > 0$) then the expected utility must be strictly decreasing in the reputation cost θ. Moreover, given the discrete reports and linear-in-parameter reputation cost, the expected utility is a continuous, strictly-decreasing, piecewise-linear function.

The expected utility in the baseline $\mathrm{EX}_\emptyset$ environment using the calibrated parameters is shown as the black line in Figure 2. By way of comparison, a clear reference outcome is complete honesty by all participants—where this reference outcome is the same in all treatments. The expected utility under complete honesty in the LC+Rep model is a constant for all types, $U_0 = 1/2$, the probability of winning in a fair competition between two die rolls, with zero incurred lying or reputation costs. Relative to complete honesty, the expected utility for participants with near-zero reputation cost in the EX environments starts above U_0, as these types can guarantee themselves a payoff of U_0 with honesty. However, low-θ participants can do strictly better by misreporting low rolls as nines. For example, for $\theta = 0$, a 0-roll honestly reported yields a payoff of $1/20$, while reporting a 9 yields $19/20 - C$.

Without a change in the report policy $\xi^\star(\omega, \theta)$, the expected utility is strictly decreasing in θ as the incurred reputation cost is linear in θ. However, as θ increases further, the reporting policy will also adapt, where those with higher reputation costs will move away from reports associated with lying (9s) to other high payoff reports (7s and 8s). The net effect though is still a decrease in expected utility in θ. Importantly to our subsequent analyses, for high-enough θ, the expected utility can be pushed substantially below U_0. The reason behind this is that even with perfectly honest reporting, reputation costs will still be incurred whenever participants report a high roll (even if truthful). In fact, as shown in Figure 2, for high values of θ, expected utility is heavily reduced, either through reputation costs incurred from honest reports, or, for high enough reputation costs, from a downwards misreport (for example, a 4 instead of a 9).

The above discussion of the utility effects in the baseline setting—and the strictly decreasing response—points to the clearest effect from the presence of an outside option p. Participants with the highest possible reputation cost, $\theta = \kappa$, get an expected utility of $U_{\mathrm{EX}_\emptyset}(\kappa)$ in the competitive task. If the outside option p is lower than this level, then there is no effect, and behavior in the presence of the outside option is identical to the baseline. Indeed, under the calibrated model this is the predicted effect when $p = p_L = 1/4$. The predicted reports for EX_L under the calibrated model are identical to $\mathrm{EX}_\emptyset$, with complete entry predicted.

In contrast, if the outside option p exceeds $U_{\mathrm{EX}_\emptyset}(\kappa)$, then the behavior in $\mathrm{EX}_\emptyset$ can no longer be an equilibrium as high-θ types will opt out. Because the expected utility in the competitive task is strictly decreasing in θ, the equilibrium with exit is characterized by a critical reputation-cost $\kappa^\star < \kappa$. Types with reputation cost below this enter, while those with costs above exit. As such the reputation-cost distribution conditional on entry is a uniform draw over $[0, \kappa^\star]$, where the equilibrium condition is that the critical type $\kappa^\star$'s expected utility in the competitive task is exactly equal to the outside option:

$$\mathbb{E}_{\omega} u\left(\kappa^{\star}, \xi^{\star}(\kappa^{\star}, \theta); \Lambda^{\star}, \pi\right) = p.$$

A clear prediction from the model on the extensive margin is that entry is decreasing in the outside option.[7] In particular, while the calibrated model predicts no difference between $\mathrm{EX}_{\emptyset}$ and EX_L with both settings exhibiting complete entry, there are substantial effects as we increase the outside option still further in EX_H, where the model predicts entry to drop to 25 percent.

While the focus above is on the extensive margin, selection also has effects on the intensive margin. In equilibrium, though, there are two effects moving in offsetting directions. High-reputation–cost types are the most-honest participants in the baseline setting, and their exit leads to an increase in the proportion of liars at high reports. But the increase to $\Lambda^{\star}(x)$ for high reports leads to an offsetting shift towards greater honesty for the remaining medium-θ participants. The report distribution under the calibrated parameters for EX_H illustrated in the bottom right of Figure 1 indicates an increases to the fraction of both maximal and just-below maximal reports for EX_H, but is otherwise fairly similar to the baseline. The expected utility effects for EX_H are illustrated through a comparison of the black and gray lines in Figure 2. All θ-types above the selection threshold opt out, while those that remain get strictly worse outcomes than they would have under no-exit. This utility reduction comes about because the types that exit were providing a positive externality in the baseline environment by reducing $\Lambda^{\star}(x)$ for high-x with their honesty.

Having examined the effects of selection in isolation, we now turn to the predicted effects of endogenizing the prizes through competition, starting from the effect in $\mathrm{EN}_{\emptyset}$ where we do not allow for exit. Where the matched roll was an exogenous and honest response ω_j, the expected reward $\pi(x)$ from a report x is

$$\pi(x) = \Pr\left\{x > \omega_j\right\} + \frac{1}{2}\Pr\left\{x = \omega_j\right\}.$$

However, when we endogenize the matched roll, the prize structure in competition becomes an equilibrium object, determined by the overall reporting distribution $\xi^{\star}$:

$$\pi^{\star}(x; \xi^{\star}) = \Pr\left\{x > \xi^{\star}(\omega, \theta)\right\} + \frac{1}{2}\Pr\left\{x = \xi^{\star}(\omega, \theta)\right\}.$$

While the exogenous-match expected prize $\pi(x)$ is linearly increasing from a low of $1/20$ to a high of $19/20$, the competitive prize structure exhibits two effects: (i) a downwards level

[7] While the direction of this comparative static is obvious, we note that within our parameterization, this hypothesis only has bite for $p < 1/2$ due to the reputation costs. Without them agents in the EX_L and EX_H treatments can ensure themselves a payoff of $1/2$ with perfectly honest reports.

shift in the prizes at all reports resulting in $\pi^\star(x) < \pi(x)$ for all x, as the competitive report distribution stochastically dominates the honest report distribution; and (ii) increasing returns to the very highest reports, so convexity in $\pi^\star(x)$.[8]

The effects on expected utility as we introduce competition are clear: Every type θ does strictly worse under competition, due to both the level effect from the reduced prizes and from the greater convexity. This is illustrated in Figure 2 where the red line ($U_{EN_\emptyset}$) indicates the expected utility to each type θ in the $\mathrm{EN}_\emptyset$ treatment. While the outcomes for all agents are shifted downwards, the predicted report distributions illustrated in Figure 1 indicate fairly muted effects (comparing the top-left and bottom-left distributions). On its own, competition across the prizes has only a limited effect in terms of degrading honesty.[9]

Finally, we consider what happens when we have both selection and competition. As illustrated by the expected utility under $\mathrm{EN}_\emptyset$ in Figure 2, it is possible that low outside options have no effect, so long as the expected outcome of the highest reputation-cost type $\theta = \kappa$ exceeds the outside option. However, even a small degree of exit by high-θ types can now lead to a substantially different equilibrium, where the next stable equilibrium outcome under the low outside option is illustrated by the blue line (U_{EN_L}). The reason for this is that as high-reputation-cost types exit, the prize function $\pi^\star()$ both shifts downward and becomes more convex (while reputation costs also increase). These effects serve to push the expected utility curve downwards for all remaining participants, leading to further exit, and so on. As such, even for a relatively low outside option, the effect of selection in concert with competitive outcomes is dramatic, with only 20 percent entry predicted. In terms of the competitive task reports, the increased convexity reduces the reward for reputationally-motivated partial lies (eights and sevens) forcing agents to essentially either lie maximally or report honestly. The top-middle panel in Figure 1 illustrates the report prediction corresponding to this equilibrium for EN_L under the literature calibration, with a sharp increase in maximal reporting and the absence of partial lies, though still with some honest reporting.

[8]While $\Lambda^\star(x)$ must be weakly increasing in equilibrium, for any non-maximal report x with $\Lambda^\star(x) > 0$, equilibrium requires that $\Lambda^\star(z) - \Lambda^\star(x) > 0$ for all $z > x$, otherwise all agents lying at x would instead report z and do strictly better. This implies that likelihood of the competing report y, $\Pr\left\{\xi^\star(\omega,\theta) = y\right\}$, is strictly increasing for $y \geq x$ (giving us stochastic dominance over the uniform) and that the marginal return of a higher report $\pi^\star(z+1) - \pi^\star(z)$ is strictly increasing (convexity).

[9]Similar to endogeneity in the population, the fixed point solutions here allow for the potential for multiple equilibria. Our focus here is on outlining intuition for the theory, where we characterize the most-honest stable equilibria under the calibrated model.

While it is possible to find non-boundary equilibria for small to moderate outside options, with large outside options the only equilibrium is complete dishonesty within the competitive task. As the selected effect increases, the expected outcome from a *maximal* report declines towards 1/2. Even with no incurred reputation costs, the direct lying costs are simply too large for any type to enter. This is the case at the calibrated payoffs for the EN_H setting, where the only possible outcome is complete dishonesty.[10]

In Table 1b we provide quantitative predictions for our treatments under the calibrated model.[11] Here we indicate the predicted entry rate by treatment, the average expected report, the fraction of reports that are maximal, and that are minimal. While the table does help contextualize the results, the exact quantitative predictions are sensitive to the parameter values used. However, in a qualitative sense the theory makes the following qualitative predictions in response to our three main questions.

Hypothesis 1 (Competition only)**.** *Competition over prizes makes participants worse off, but has minimal effect on the distribution of reports.*

Hypothesis 2 (Selection only)**.** *An increase in the size of the outside option (expected payoff in the fixed task) decreases the frequency of entry to the competitive task, but has minimal effect on the distribution of reports.*

Hypothesis 3 (Competition and selection)**.** *The combination of competition and selection results in greater dishonesty and lower entry relative to the comparable treatments with either effect in isolation.*

Hypothesis 1 outlines an effective null results when we compare the report distributions between $\mathrm{EX}_\emptyset$ and $\mathrm{EN}_\emptyset$. While the hypothesis also outlines a reduction in utility, this is not directly observable solely through payoffs as non-observable reputation costs will also be a part of this.

In contrast Hypothesis 2 makes two predictions that are both testable in the data. As we introduce a small outside option, the hypothesis is that there is no effect on the report distributions as we move from $\mathrm{EX}_\emptyset$ to EX_L, and no effect on entry (with full entry predicted in both treatments). However, as the outside option increases in value, the prediction is

[10] To avoid cycles, this conclusion relies on there being a small positive mass of agents ϵ who have neither lying costs nor reputation costs. Allowing for type heterogeneity over both θ and C yields a similar outcome, with some positive measure close to the zero costs point, though with much greater complexity in the reporting policy.

[11] Other than EN_L we indicate the most-honest equilibria. For EN_L an identical outcome to $\mathrm{EN}_\emptyset$ exists as a stable outcome, where the table indicates the next-most-honest stable outcome.

for large reduction in entry into the competitive task as we move from EX_L to EX_H (with a slight decrease in dishonesty, primarily through a 66 percent increase in maximal reports, cf Table 1.

Finally, Hypothesis 3 makes a prediction on the interactive effect. Relative to both $\text{EX}_\emptyset$ and EX_L, the theory predicts that competition and selection will lead to a large reduction in entry in EN_L, as well as a substantial decrease in honesty (a 133 percent increase in maximal reports). The predicted effect becomes complete for the EN_H treatment, where the theory predicts that only those with no dishonesty costs remain (so a substantial decrease in entry and honesty for EN_H relative to any other treatment).[12]

3. Results

We begin by looking at the aggregate results in Table 2, presenting both averages and qualitative tests across treatments. For each treatment the table indicates: (i) the *entry rate* into the competitive task; (ii) the average *fixed-task report*; (iii) the average *competitive-task report*, and the proportion of competitive task reports that are maximal ($x = 9$).

Below each treatment-level average we provide participant-clustered standard errors.[13] Using the average responses we estimate pairwise comparative statics inferences for the treatments. Drawing a clear line in the sand, we use the notation $X \succ Y$ to indicate that the average in treatment X is greater than in treatment Y with $p \leq 0.001$ on a two-sided test of equality. Further, we use the notation $X \sim Y$ to indicate failure to reject equality with $p > 0.1$, allowing the presented empirical order to potentially be incomplete for intermediate significance levels. In the last two columns of Table 2 we present the outcome of Wald tests examining the joint effects across the separate design dimensions.[14] In the EX=EN column we compare the treatment averages for the exogenous prize treatments (EX) against the comparable endogenous prize treatment (EN). Similarly, in the $L = H$ (and $\emptyset = L = H$) column we compare the effects between the exogenous and endogenous treatment pair for the *Low* and *High* outside options (also *None* where it is non-mechanical).

[12] We had not preregistered these hypotheses, as the planning for these experiments predated the norm for preregistration. However, the very detailed *quantitative* predictions given in Table 1 are pinned down by the preferences and parameters imported from Abeler et al. (2019), which in some sense serves as much more-restrictive pre-registration, which hopefully allays some reader concerns here.

[13] All results are from a joint regression of the relevant outcome variable on a set of orthogonal treatment dummies.

[14] Here we present p-values as each entry is a single test; however, as can be seen, the incomplete inference order $\succ$ can be similarly employed.

TABLE 2. Summary Statistics and Treatment Differences

Variable	Exogenous competition				Endogenous competition				Joint Tests	
	None	Low	High	Order	None	Low	High	Order	***EN*=*EX***	$\boldsymbol{L = H}$
Entry Rate	1	0.864 (0.030)	0.538 (0.036)	$\mathrm{EX}_L > \mathrm{EX}_H$	1	0.649 (0.030)	0.404 (0.031)	$\mathrm{EN}_L > \mathrm{EN}_H$	$p < 0.001$	$p < 0.001$
	Fixed Task									
Avg. Report	–	4.52 (0.24)	4.52 (0.14)	$\mathrm{EX}_L \sim \mathrm{EX}_H$	–	4.72 (0.09)	4.58 (0.07)	$\mathrm{EN}_L \sim \mathrm{EN}_H$	$p = 0.690$	$p = 0.452$
	Competitive Task								***EX*=*EN***	$\boldsymbol{\emptyset = L = H}$
Avg. Report	5.83 (0.28)	5.76 (0.28)	6.01 (0.22)	$\mathrm{EX}_H \sim \mathrm{EX}_L \sim \mathrm{EX}_\emptyset$	5.93 (0.16)	7.68 (0.19)	7.29 (0.24)	$\mathrm{EN}_L \sim \mathrm{EN}_H > \mathrm{EN}_\emptyset$	$p < 0.001$	$p < 0.001$
Max Reports	0.240 (0.048)	0.206 (0.045)	0.181 (0.017)	$\mathrm{EX}_\emptyset \sim \mathrm{EX}_L \sim \mathrm{EX}_H$	0.243 (0.028)	0.639 (0.039)	0.540 (0.054)	$\mathrm{EN}_L \sim \mathrm{EN}_H > \mathrm{EN}_\emptyset$	$p < 0.001$	$p < 0.001$
Observations	720	720	720		2,160	2,160	2,160			
Participants	24	24	24		72	72	72			

Note: Raw means presented for each treatment, with participant-clustered standard errors in parentheses recovered from a linear regression. Comparative statics are obtained from t-tests for equality of coefficients, where $>$ indicates probabilistic rejection for every p above 0.001 and $\sim$ indicates failure to reject at $p = 0.1$ and below. Joint tests indicate the p-values from Wald tests over: (i) EN vs. EX column provides a joint test of $(\mathrm{EX}_L, \mathrm{EX}_H) = (\mathrm{EN}_L, \mathrm{EN}_H)$ for the entry and fixed task reports and $(\mathrm{EX}_\emptyset, \mathrm{EX}_L, \mathrm{EX}_H) = (\mathrm{EN}_\emptyset, \mathrm{EN}_L, \mathrm{EN}_H)$ for the competitive task; (ii) $\boldsymbol{L = H}$ and $\boldsymbol{\emptyset = L = H}$ are joint tests of $(\mathrm{EX}_L, \mathrm{EN}_L) = (\mathrm{EX}_H, \mathrm{EN}_H)$ and $(\mathrm{EX}_\emptyset, \mathrm{EN}_\emptyset) = (\mathrm{EX}_L, \mathrm{EN}_L) = (\mathrm{EX}_H, \mathrm{EN}_H)$, respectively.

Using Table 2 as reference, we summarize the top-line results, and outline the evidence for each. First, in terms of entry into the competitive task we find that:

Result 1. *Task choice responds to: (i) The size of the outside option, with significantly less entry with larger outside options, even when the outside-option payoff is dominated by the payoff from honesty in the competitive task. (ii) The nature of the competition, with significantly less entry when competition is endogenous.*

The experimental results indicate a clear ordering for competitive-task entry across the low and high outside options. This is true for the EX and EN treatments separately, as well as jointly across treatment pairs (all tests reject a null of equality with at least 99.9 percent confidence). The observed entry rates indicate less entry the greater the opportunity cost of doing so, consistent with the comparative statics from the LC+Rep model.

Observed competitive-task entry in the EX_L treatment is substantial at 86 percent, falling to 54 percent in EX_H when we increase the outside option from $p_L = 0.25$ to $p_H = 0.45$. While the *direction* of the outside-option effect is unsurprising, the fact that many participants do opt out does provide useful information. When the prizes are exogenously fixed, participants can guarantee themselves a chance of winning of 1/2 by entering and reporting honestly. As such, without the reputation-cost component in the preference, the competitive-task would stochastically dominate the outside option in both the EX_L and EX_H treatments. That some participants opt out in these treatments implies that close to half of the participants are willing to take a \$0.50 loss in order to avoid the competitive task, while approximately one-in-seven are willing to take a \$2.50 expected loss. As such, the entry results in the EX treatments clearly identify the presence of a reputation cost that is separate from the realized honesty, matching a prediction from the LC+Rep model.

The size of the outside option generates a similar entry effect when the prizes are determined competitively. The entry rate of 65 percent in EN_L is significantly greater than the 40 percent rate found in EN_H, although relative to the comparable EX treatments, the entry rates in EN match the qualitative theoretical prediction, with significantly reduced entry rates. However, the results here are distinct from the quantitative predictions from the calibrated model.

Moving on from the entry decisions, we turn to reporting behavior. Without an incentive to exaggerate the roll, the fixed-task reports in the outside option are essentially indistinguishable from fully honest reporting. Table 2 indicates that fixed task reports in all

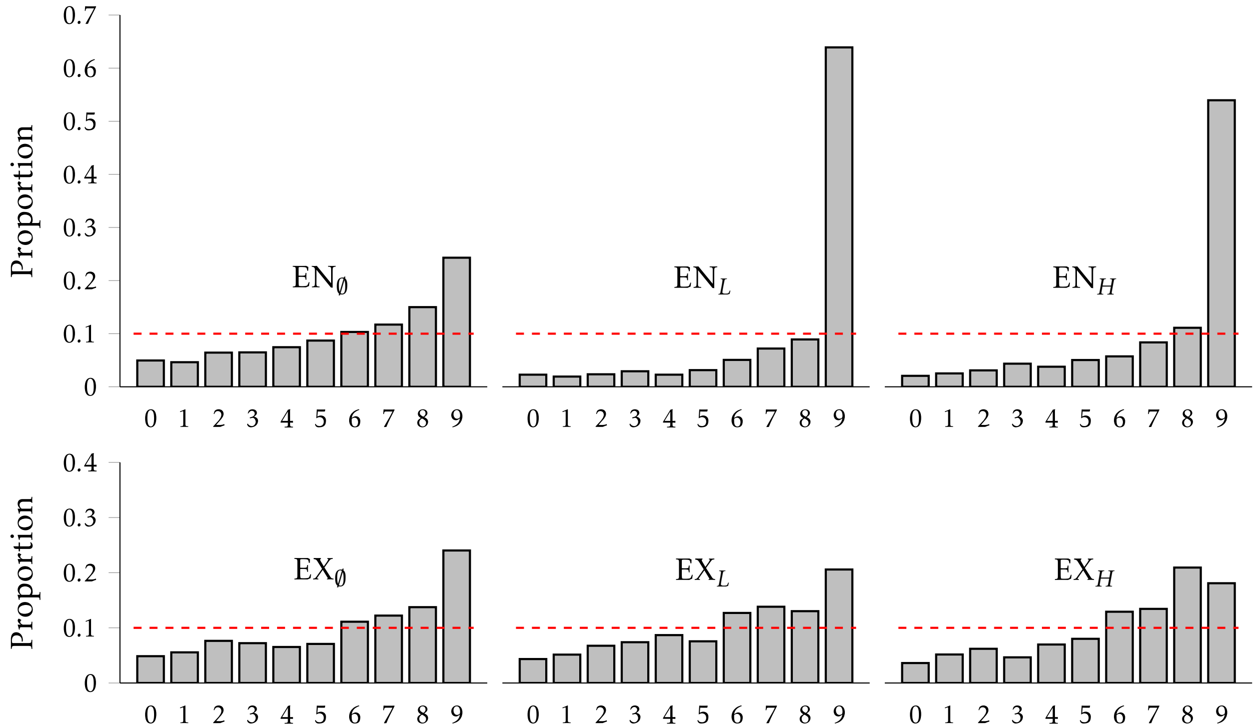

Figure 3. Observed Reports

Note: Histograms show unconditional probabilities of each die roll 0 to 9. The dotted reference line shows the true data-generating process of a 1/10 probability.

treatments are centered on 4.5, the expected outcome from a ten-sided die with faces labeled 0 to 9. A joint test across the four outside option treatments finds no significant differences.[15] The results for the means are echoed in the distributions of reports in the fixed task (see Figure A.1 in the appendix for histograms).[16]

We now turn to the results for the core outcome: the competitive-task reports. Table 2 provides two measures: the average report across the competitive task and the fraction of reports that are maximal. The main findings on the competitive-task reports can be summarized as follows:

Result 2 (Competitive-Task Honesty)**.** *Reports in the competitive task indicate:*

[15]At the treatment level, we can reject the hypothesis that the average report is 4.5 in the EN_L treatment with $p = 0.014$. However, this is due to a single participant with an average die-roll report of 6.3 across 27 fixed-task choices. Excluding this participant, the average in EN_L falls to 4.53, and we cannot reject the true average of 4.5.

[16]While the means indicate no significant differences, a Pearson chi-squared test on the pooled treatment response does reject the die's discrete uniform distribution with $p = 0.011$. The reason is a significant under-reporting of zeros. Pooling all fixed-task reports, the likelihood of observing 197 or fewer 0s from 2,475 total rolls is approximately one-in-3,580. This pattern in under-reporting zeros shows up in all four treatments.

(1) In the EX treatments, reports are significantly greater than the expected die roll, but they are not fully dishonest. The behavior mirrors the standard result in the literature, despite our task involving repetition.

(2) Reports in $EN_{\emptyset}$ treatments are similar to those in the EX treatments, where endogeneity of competition by itself does not significantly increase dishonesty.

(3) When there is both endogenous competition and task selection, behavior becomes significantly more dishonest, though there are no substantial differences on average between EN_L and EN_H.

Evidence for the above results are provided in the treatment-level comparisons in Table 2, and histograms of competitive task reports in Figure 3. The first two components of Result 2 indicate that, despite the substantial repetition of the die-rolling task in our setting, the aggregate report distributions in the EX and $\text{EN}_{\emptyset}$ treatments match the stylized result in the literature. Moreover, none of the competitive-report metrics in Table 2 show any significant difference between the four treatments. As such, in isolation neither endogenous competition nor task-selection generate differences relative to the standard environment ($\text{EX}_{\emptyset}$).

We see no difference in average reports between $\text{EX}_{\emptyset}$, EX_L, EX_H, and $\text{EN}_{\emptyset}$, with $p \geq 0.476$ in all pairwise comparisons, although average reports are significantly different from honest reporting ($p < 0.001$ for all treatments), with 23 percent of reports being maximal. This 13 percentage-point excess of maximal reports indicates some lying, but is far from the payoff-maximizing prediction that all reports should be maximal. In particular, for these four treatments we can reject a maximal reports excess of 20 percentage points or more with $p < 0.001$.

Neither endogenous competition nor the presence of an outside option generate high levels of dishonesty in isolation; however, the final component of Result 2 indicates that the *combination* of both endogenous competition and selection dramatically increases dishonesty. Evidence for this claim is most apparent when comparing the histograms in Figure 3 where only one (or neither) of the effects is present ($\text{EX}_{\emptyset}$, EX_L, EX_H, and $\text{EN}_{\emptyset}$) to the histograms where both forces are present (EN_L, EN_H). Sixty percent of competitive task reports are maximal in the pooled EN_L and EN_H data. The ensuing 50 percentage point excess of reported nines in these treatments represent almost quadruple the dishonesty rate from the other four treatments. Both the rate of maximal reporting and the average report in Table 2, have the treatment ordering

$$\text{EN}_L \sim \text{EN}_H \succ \text{EN}_{\emptyset} \sim \text{EX}_{\emptyset} \sim \text{EX}_L \sim \text{EX}_H;$$

where $\sim$ indicates failure to reject at $p = 0.10$ and $\succ$ indicates rejection of equality at $p < 0.001$.[17]

We decompose the relative effects of each treatment dimension on honesty through a regression. Standardizing each subject's competitive-task report $\overline{X}_i$ to an "implied honesty" measure $\hat{\lambda}_i := 2 - \overline{X}_i/4.5$, we regress the implied honesty on dummy variables representing the possibility of selection via an outside option (*SEL*), the presence of endogenous reports in the competitive task (*EN*), and the interaction of the two effects.[18] The estimated econometric equation (with participant-clustered standard errors in parentheses) is given by:

$$\hat{\lambda} = \underset{(0.062)}{0.706} - \underset{(0.076)}{0.007} \cdot \mathrm{SEL} - \underset{(0.072)}{0.023} \cdot \mathrm{EN} - \underset{(0.090)}{0.349} \cdot \mathrm{SEL}{\times}\mathrm{EN}$$

The results indicate that neither the selection nor competition channels on their own lead to any statistically significant difference with the baseline of 70 percent honesty. However, the interaction of endogenous selection with competition does produce a significant effect, reducing the effective honesty rate by a half.

Finally, while we do find a strong effect on honesty from the interaction of selection and competition, when we compare the EN_L and EN_H treatments, we do not find any effect from the *size* of the outside option. Indeed, the theoretical prediction is that the upper-bound honesty sustainable in equilibrium should fall as the size of the outside option grows. Though insignificant ($p = 0.206$ for reports and $p = 0.140$ for maximal reports) the results instead point in the opposite direction from the prediction, with greater observed honesty in EN_H. We come back to try and explain this facet of the data in the next section when we examine group-level dynamics.

Discussion: Participant and Group Heterogeneity. To understand why the interaction of selection and competition has such a large effect, we analyze behavior at the participant and group levels. Looking at the participant sub-sample with at least 5 competitive-task observations we divide the data into the treatments with both endogeneous market features (EN_L and EN_H) and those with one or neither (the three EX treatments and $\mathrm{EN}_\emptyset$). In Figure 4 we present density estimates across the participant-level honesty alongside a reference distribution showing what fully honest play would look like, using the same

[17] In realization the data comparative static here is fully complete and transitive.

[18] The range of $\hat{\lambda}$ is 0–2, where maximal reports yield a value of $\hat{\lambda} = 0$, indicating fully dishonest behavior, while $\hat{\lambda} = 1$ indicates honest behavior on average. Values of $\hat{\lambda} > 1$ indicate "supra-honest" behavior, in which a participant's average reports are less than would have been expected by chance.

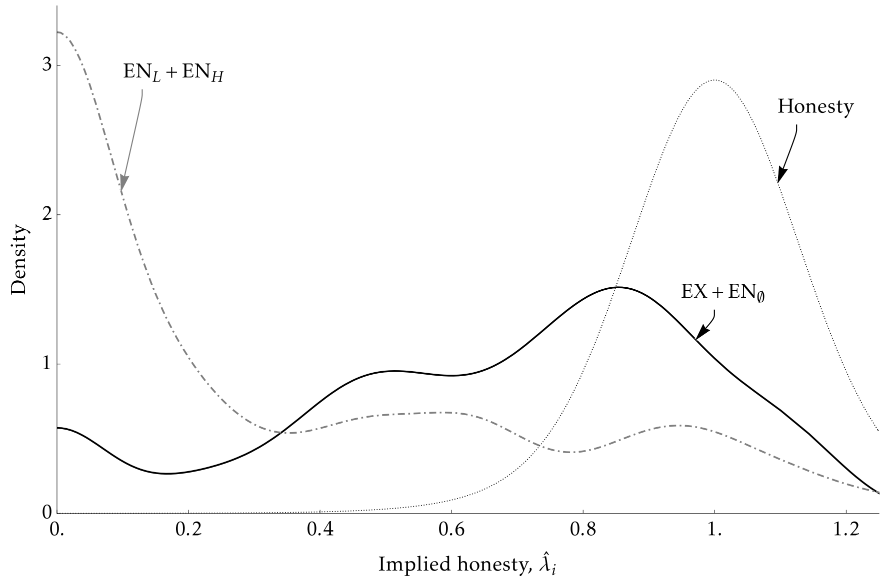

FIGURE 4. Participant-level honesty in the competitive task

Note: Figure indicates kernel-smoothed density estimates of subject-level honesty. Honesty reference line indicates expected distribution under complete honesty due to random sampling fixing subject-level engagement in competitive task. Black line pools the EX and $\text{EN}_\emptyset$ treatments (144 subjects). Dot-dashed line pools the EN_L and EN_H treatments (144 subjects).

number of competitive-task observations per participant.[19] Comparing the individual-level response distributions illustrates that the core differences in the distributions are driven by a large subset of participants in EN_L and EN_H that behave completely dishonestly, choosing near-maximal reports in each round. Forty-six percent of the participants in these two treatments have an implied honesty rate below 0.22 (an average competitive-task report in excess of eight) where the median honesty rate is 0.27. In contrast, for the other four treatments we find much greater participant-level honesty, where the median participant is only slightly dishonest at 0.76, and where only eight percent of individuals have an implied honesty below 0.22.

Examining the participant-level distributions, a natural question is whether individual participants have become more dishonest in the EN_L and EN_H treatments (an intensive margin shift), or instead whether the competitive-task population has simply become more selected (an extensive margin shift), with the honest participants opting out. To partially address this, we can use the report behavior from the $\text{EN}_\emptyset$ treatment, where there is no selection, to understand what the distribution would look like were we to remove the most-honest participants.

[19] While the reference distribution has a clear mode at 1, it illustrates the inherent variability we should expect under honest play due to the die-roll realizations.

Using participant-level honesty rates in $EN_{\emptyset}$ we form pseudo-matching groups of six participants, where we then remove the most-honest individuals within each simulated group, holding constant the reporting behavior of the remaining participants.[20] While the average honesty with all six group members is simply the treatment average of 0.68, removing the most-honest group-member (as if they selected into the fixed task) reduces the group average to 0.62. Similarly, removing the second-, third- and fourth-most-honest participants reduces the implied honesty of the remaining group to 0.55, 0.47 and 0.38 respectively. Using linear interpolation for exit rates in EN_L and EN_H of 2.1 and 3.6, respectively, leads to predicted pure-selection honesty rates of 0.541 and 0.419. These results show that, while it is plausible that the observed EN_H honesty rate of 0.379 could possibly be driven by pure-selection ($p = 0.438$ on a point test), the observed rate of 0.294 in EN_L is substantially smaller than the predicted level ($p < 0.001$). Using the averages to get a sense of the magnitudes, the results here suggest that 13 percent of the dishonesty in EN_H is driven by an intensive margin equilibrium effect, and 64 percent in EN_L.

Relative to the theoretical predictions from the population-level model, a puzzle exists over the relatively greater dishonesty in EN_L than EN_H, and the very different quantitative conclusions over the intensive margin effects. One potential explanation is simply that there are multiple equilibria here; where our theoretical predictions indicate the most-honest outcome. That is, it could be that in EN_L behavior has simply coordinated on a different equilibrium outcome, the fully uninformative outcome. However, that answer is unsatisfying, as it does not address the excess honesty in EN_H, where the unique prediction for this treatment is complete dishonesty.

While not part of our original hypotheses, another explanation emerges when we consider the heavy-lifting that the “population level” approach is doing for our predictions. Instead of an infinite rematching population with uniformly distributed reputation costs, our experiments use matching groups of six individuals. With multiple rounds of rematching, it becomes possible for subjects to learn the realized distribution of reputation costs within their group. If a group of individuals with very high reputation costs were matched together repeatedly, it would be possible for them to theoretically support higher honesty, even without folk-theorem-like forward arguments. This argument becomes even stronger once you consider the idea that with substantial exit, the effective rematching group becomes even smaller still. Some sense for the validity of this argument can be assessed by looking at heterogeneity at the group level.

[20] All simulations use 10,000 draws to construct expectations.

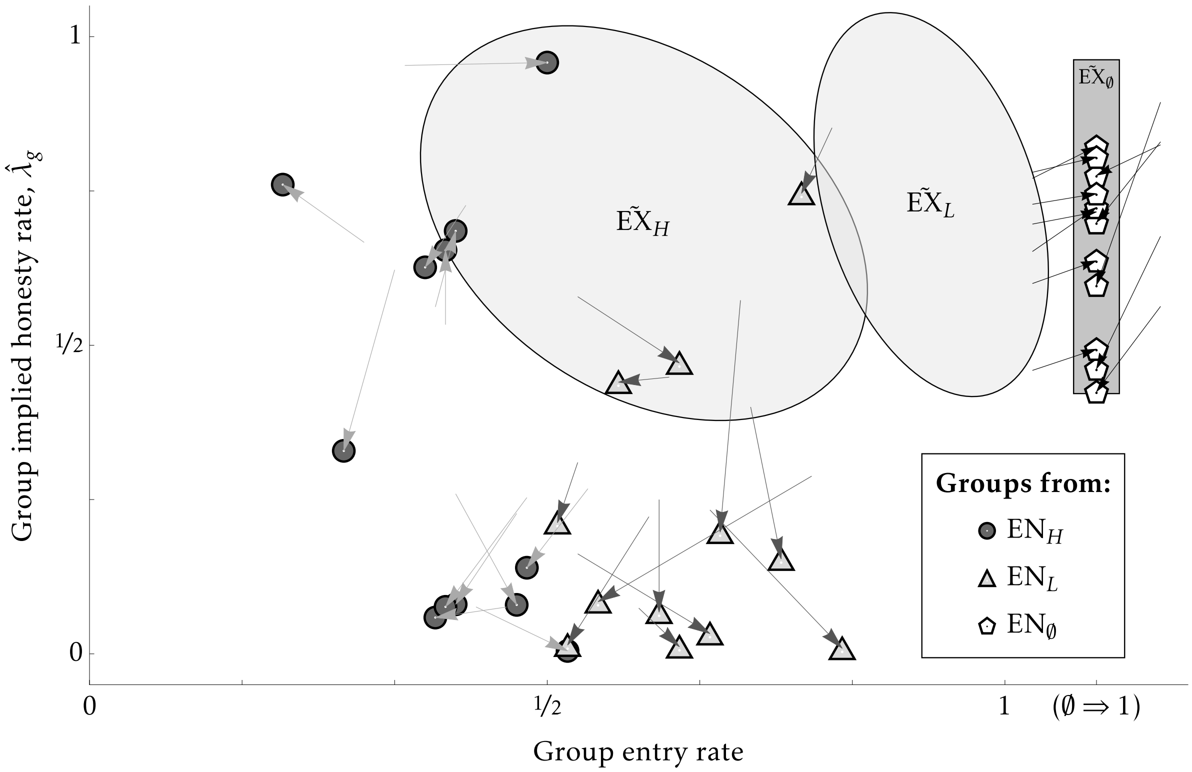

FIGURE 5. Group Variation in Task and Honesty

Note: Each point indicates the average entry rate (horizontal axis) and honesty (vertical axis) by 6-person matching group in the second half of sessions (rounds 16–30). Arrows indicate the movement for the same group from the first half (rounds 1–15). EX treatments have fixed entry and indicated to the left of figure. Shaded regions represent 95 percent coverage region for recombinant groups of six formed from subjects in our EX individual treatments.

In Figure 5 we illustrate group-level heterogeneity, as well as some of the within-group dynamics. On the horizontal axis we indicate the average group-level entry rate into the competitive task (the extensive margin), while the vertical axis depicts the average group-level implied honesty in the competitive task, $\hat{\lambda}_g$ (the intensive margin). Each plotted point indicates the average extensive and intensive margin behaviors in the last half of the session (rounds 16–30) for the 12 matching groups in each EN treatment. White pentagons indicate the group averages in $EN_\emptyset$, lighter gray triangles for EN_L and dark gray circles for EN_H. The arrow pointing to each plotted point indicate the within-session dynamics, where the arrow's origin point indicates the same group's averages in the first half of the session (rounds 1–15).[21] Finally, to provide a comparable group-level statistics

[21]Entry for $EN_\emptyset$ is full by construction, so to make this clear we shift these groups averages to a separate area on the right to indicate this feature of the data is mechanical. Moreover, because entry is fixed, for clarity in the $EN_\emptyset$ treatment the arrows move in from the right or left depending on whether the group's honesty increased or decreased.

for the behavior in our Exogenous treatments (which are decision problems), we recombinantly form groups of six from the participants in these individual-decision treatments, and graph the 95 percent confidence intervals over these pseudo-groups' averages as the two shaded ellipses.

Inspecting the group-level behavior, a first conclusion is there is too much variability in the EN_H treatment, with a more bi-modal distribution in the vertical axis. For the 12 circles representing EN_H groups, the average implied honesty across all groups is 0.40, but with a standard deviation across groups of 0.31. To show this is too large, we can use simulate 12 groups with pure selection from the $\mathrm{EN}_\emptyset$ population. Using this method, we can generate a similar expected group-level average of 0.42, but with much smaller expected standard deviation across the groups of 0.18. Generating the sampling distribution for the standard deviation across the simulated groups, we find a 99.9-percent confidence interval of $[0.07, 0.29]$, so we can conclude that variability across the EN_H groups is very unlikely to be a chance realization. Intuitively, pure selection would lead to a more-centered distribution of the group averages. In contrast, the actual EN_H groups illustrated in Figure 5 have either high honesty levels (comparable to the average in $\mathrm{EN}_\emptyset$) or very low honesty close to zero. As such, while we can reject a pure selection in EN_L due to the absolute levels of the observed honesty, we can still reject pure selection in EN_H due to too much variability.[22] To have both a lower average, and greater variability requires some interaction between participant honesty rates within each group. In particular, for the six groups with almost complete dishonesty, participants must be responding to the dishonesty of others by increasing their own dishonesty on the intensive margin.

A second piece of evidence here is shown in the dynamics. Each arrow shows the movement in the group-level averages across the sessions. Pure extensive-margin selection would show up with arrows with large movements from right to left. But instead the figure indicates minimal horizontal movements. Instead, selection is more substantive in the vertical direction, indicating shifts in the honesty across the session. Groups that have lower honesty in the session's first half tend to have even lower honesty in the second half of the session. Moreover, for the EN_H treatments, we actually see increases across the session for many of the groups that were most honest in the first half.

The theoretical prediction that increased selection pressure creates complete dishonesty in the EN_H treatment is driven by each agent being a tiny element of a large population.

[22] The pure-selection model for EN_L predicts an expected sample-standard-deviation across 12 groups of 0.15 with a 98-percent confidence interval of $[0.09, 0.23]$. The sample group-level standard deviation in EN_L is instead 0.22. So, while we can reject pure selection here based on the mean, the same finding of too much heterogeneity still holds at 98 percent confidence.

While our average results certainly indicate far more dishonesty in the EN_Htreatment than the baseline $\mathrm{EX}_\emptyset$ setting with no market forces, the group-level heterogeneity here indicates how the conclusions can be different with small rematching populations. As Figure 5 makes clear on the horizontal axis, the smaller the effective matching group (the further to the left on the figure), the greater the heterogeneity we observe in the competitive-task honesty.

While the effects of small groups is outside of what we can predict with a population-level model—which is already analytically cumbersome—the reversal of the theoretical comparative static behavior between our EN_L and EN_H treatments speaks to a more heartening idea. In repeated settings, even with both selection and competition, partial honesty can be a stable outcome in small enough populations. This idea can be interpreted to be something like "small-town values," or that with elite enough selection in domains such as politics, more-efficient norms might be supported. While this result does offer a glimmer of hope, the population-level theoretical predictions also point to the precariousness of these norms with both types of market selection. Someone with fewer scruples comes to town, and the norms can rapidly change: forcing out those with strong moral qualms, and increasing the incentives in a race to the bottom.

4. Conclusion

Our results show that competition has the potential to significantly degrade honesty, but that effect depends heavily upon the structure of the competition. Our exogenous treatments with an outside option demonstrate that selection alone does not markedly increase dishonest behavior, while our $\mathrm{EN}_\emptyset$ environment indicates that without selection, endogenous rewards can still sustain substantially honest behavior. Only once we combine the two market forces, with voluntary entry and endogenous matching, do we observe large increases in dishonesty.

Our findings and analysis of the lying and reputation cost models of dishonesty suggest subtlety in interpreting the effect of market forces on honesty. While report competition between individuals does lower payoffs relative to competition with an honest robot, it does not fully degrade dishonesty. The continuing presence and honesty of the most reputation-sensitive types helps to maintain an incentive for partial honesty by those with more-moderate reputation costs. Likewise, our exogenous treatments that allow selection out of the competitive task only have a moderate effect on dishonesty. Despite substantial observed selection, the logic of reputation costs acts to inhibit extreme dishonesty by those remaining. However, once both margins are present, dishonesty does becomes the norm. More-honest types exit the competition rather than incur the psychic costs of lying, resulting in higher reputational costs for those who remain and lower chances of winning, driving out those with moderate concern for reputation. This process continues until the only types remaining in the competition are those with the least concern for reputation, those that will do not hesitate to lie at the extremes.

In aggregate, our experimental results echo the theoretical prediction that a combination of selection and competition will have a large negative effect on honesty. However, we also observe considerable heterogeneity in outcomes at the group level. A sizeable minority of groups do manage to sustain relatively honest behavior, even in the long run. One useful avenue for future work would be to explain the dynamics that predict whether a group's long-run outcomes will devolve to dishonesty or maintain truthful behavior in equilibrium. There are also opportunities to extend our framework to other types of competitions or tasks, such as the real-effort tournaments that have been extensively studied in the past.

References

Abeler, Johannes, Daniele Nosenzo, and Collin Raymond, "Preferences for truth-telling," *Econometrica*, 2019, *87* (4), 1115–1153.

Barfort, Sebastian, Nikolaj A Harmon, Frederik Hjorth, and Asmus Leth Olsen, "Sustaining honesty in public service: The role of selection," *American Economic Journal: Economic Policy*, 2019, *11* (4), 96–123.

Belot, Michèle and Marina Schröder, "Sloppy work, lies and theft: A novel experimental design to study counterproductive behaviour," *Journal of Economic Behavior & Organization*, 2013, *93*, 233–238.

Benistant, Julien, Fabio Galeotti, and Marie Claire Villeval, "Competition, information, and the erosion of morals," *Journal of Economic Behavior & Organization*, 2022, *204*, 148–163.

Cadsby, C Bram, Fei Song, and Francis Tapon, "Are you paying your employees to cheat? An experimental investigation," *B.E. Journal of Economic Analysis & Policy*, 2010, *10* (1).

Carpenter, Jeffrey, Peter Hans Matthews, and John Schirm, "Tournaments and office politics: Evidence from a real effort experiment," *American Economic Review*, 2010, *100* (1), 504–17.

Cartwright, Edward and Matheus LC Menezes, "Cheating to win: Dishonesty and the intensity of competition," *Economics Letters*, 2014, *122* (1), 55–58.

Casella, Alessandra, Navin Kartik, Luis Sanchez, and Sébastien Turban, "Communication in context: Interpreting promises in an experiment on competition and trust," *Proceedings of the National Academy of Sciences*, 2018, *115* (5), 933–938.

Conrads, Julian, Bernd Irlenbusch, Rainer Michael Rilke, Anne Schielke, and Gari Walkowitz, "Honesty in tournaments," *Economics Letters*, 2014, *123* (1), 90–93.

Dannenberg, Astrid and Elina Khachatryan, "A comparison of individual and group behavior in a competition with cheating opportunities," *Journal of Economic Behavior & Organization*, 2020, *177*, 533–547.

d'Adda, Giovanna, Donja Darai, Nicola Pavanini, and Roberto A Weber, "Do leaders affect ethical conduct?," *Journal of the European Economic Association*, 2017, *15* (6), 1177–1213.

Faravelli, Marco, Lana Friesen, and Lata Gangadharan, "Selection, tournaments, and dishonesty," *Journal of Economic Behavior & Organization*, 2015, *110*, 160–175.

Fehrler, Sebastian, Urs Fischbacher, and Maik T Schneider, "Honesty and self-selection into cheap talk," *Economic Journal*, 2020, *130* (632), 2468–2496.

Feltovich, Nick, "The interaction between competition and unethical behaviour," *Experimental Economics*, 2019, *22* (1), 101–130.

Fischbacher, Urs, "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, 2007, *10* (2), 171–178.

______ **and Franziska Föllmi-Heusi**, "Lies in disguise—an experimental study on cheating," *Journal of the European Economic Association*, 2013, *11* (3), 525–547.

Gneezy, Uri, "Deception: The Role of Consequences," *American Economic Review*, March 2005, *95* (1), 384–394.

______**, Agne Kajackaite, and Joel Sobel**, "Lying Aversion and the Size of the Lie," *American Economic Review*, 2018, *108* (2), 419–53.

______**, Bettina Rockenbach, and Marta Serra-Garcia**, "Measuring lying aversion," *Journal of Economic Behavior & Organization*, 2013, *93*, 293–300.

Hanna, Rema and Shing-Yi Wang, "Dishonesty and selection into public service: Evidence from India," *American Economic Journal: Economic Policy*, 2017, *9* (3), 262–290.

Konrad, Kai A, Tim Lohse, and Sven A Simon, "Pecunia non olet: on the self-selection into (dis) honest earning opportunities," *Experimental Economics*, 2021, *24* (4), 1105–1130.

Mazar, Nina, On Amir, and Dan Ariely, "The dishonesty of honest people: A theory of self-concept maintenance," *Journal of Marketing Research*, 2008, *45* (6), 633–644.

Necker, Sarah and Fabian Paetzel, "The effect of losing and winning on cheating and effort in repeated competitions," *Journal of Economic Psychology*, 2023, *98*, 102655.

Olsen, Asmus Leth, Frederik Hjorth, Nikolaj Harmon, and Sebastian Barfort, "Behavioral dishonesty in the public sector," *Journal of Public Administration Research and Theory*, 2019, *29* (4), 572–590.

Rigdon, Mary L and Alexander P D'Esterre, "The effects of competition on the nature of cheating behavior," *Southern Economic Journal*, 2015, *81* (4), 1012–1024.

Schwieren, Christiane and Doris Weichselbaumer, "Does competition enhance performance or cheating? A laboratory experiment," *Journal of Economic Psychology*, 2010, *31* (3), 241–253.

Appendix A. Additional Figures and Experimental Instructions

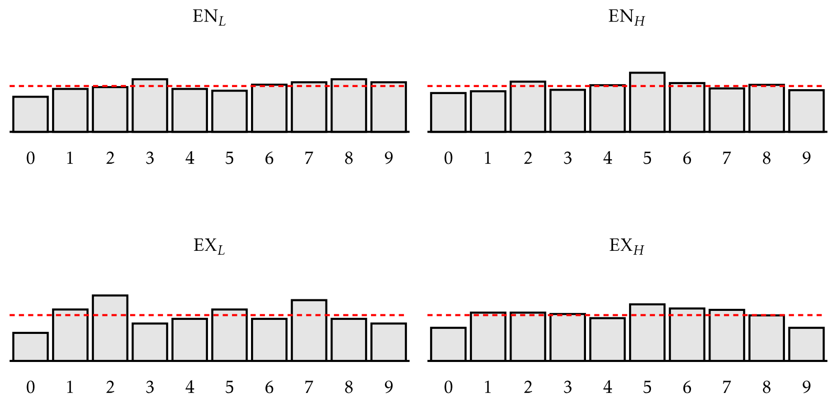

Figure A.1. Histogram of reports in Fixed Task

Appendix B. Representative Instructions

B.1. Instructions for Endogenous-High.

Instructions for Today's Experiment

Introduction. Welcome to this experiment on decision-making. Please read these instructions carefully as they explain how you earn money from the decisions you make in this experiment. There will be no talking during today's session. If you have a question, please raise your hand and an experimenter will answer your question in private.

You will remain anonymous during the experiment and after it has concluded. Only a randomly assigned ID number will identify your decisions, and any research data collected during the course of the study will only identify your decisions by that number. The only personally identifiable information that will be recorded is that you participated in this session and were paid.

Your participation in this research study is completely voluntary. Should you change your mind about participating, you can withdraw from the study at any time. Your participation in the study will remain anonymous, and no identifying records will be retained of your withdrawal from the study. However, if you do decide to withdraw after or during the study, because of the anonymity inherent to the data collection, the researchers will be unable to delete the record of your responses.

Your current and future status with the University of Pittsburgh and any other benefits for which you qualify will be the same whether you participate in this study or not. If you withdraw during the experiment you are entitled to a $6 show-up payment, where the completed study payments (between $10 and $30) are only given to subjects who finish the study.

This study is being conducted by Alistair Wilson who can be reached at alistair@pitt.edu

Preliminaries.

- On the desk in front of you is a ten-sided die. Please put your hand up now if there is not a ten-sided die at your desk.
- We will explain in a moment how you will use this die in the experiment. For now, please roll the die several times to ensure that you are happy that it is a fair die and that you know how to read the outcome.
- The sides of the die have ten numbers ranging from 0 (the lowest) to 9 (the highest), where the numbers $\underline{6}$ (six) and $\underline{9}$ (nine) are each underlined, to avoid confusing them for one another.

Description of the Experiment. The experiment will consist of 30 rounds. In each round you need to do two things:

(1) Select one of two different tasks (A or B) to participate in that round.

(2) Roll your die one time, and report the outcome.

Depending on the task you choose and the outcome of your die roll, you will receive a round payment of either $15 or $5.

The two tasks you can choose from are:

Task A: Odd/Even Matching. If you select this task you will roll your die and report the roll. After you report the result the computer draws a ball from a virtual urn that will be used to determine your earnings.

The computer urn contains one-hundred balls. Forty of the balls are labeled *EVEN*, forty balls are labeled *ODD*, fifteen balls are labeled *NEITHER*, and five balls are labeled *BOTH*.

As each ball is selected with equal probability there is therefore a 40-in-100 chance that *EVEN* is selected; a 40-in-100 chance that *ODD* is selected; a 15-in-100 chance that *NEITHER* is selected; and finally and a 5-in-100 chance that *BOTH* is selected.

Your round earnings in *Task A* are calculated as follows:

- If an *EVEN* ball is drawn you will receive $15 for the round if your roll was an even number (0, 2, 4, 6 or 8) and $5 otherwise.
- If an *ODD* ball is drawn you will receive $15 for the round if your roll is odd (1, 3, 5, 7 or 9) and $5 otherwise.
- If a *NEITHER* ball is drawn, you will receive $5 for the round regardless of your roll.
- If a *BOTH* ball is drawn, you will receive $15 for the round regardless of your roll.

Task B: High Rolling. If you select this task you will roll your die one time for the round and report the roll. After you report the result the computer then puts your roll in competition with *another* roll.

Your round earnings in *Task B* are calculated as follows:

- If your roll is higher than the competing roll you earn $15 for the round. For example if you rolled a 5 and the competing roll was between 0 and 4.
- If your roll is lower than the competing roll you earn $5 for the round. For example if you rolled a 5 and the competing roll was between 6 and 9.
- If the two rolls are the same, the computer randomly chooses one of the two rolls with equal probability (a 50-in-100 chance) to be the winner, as if flipping a fair coin. If your roll wins you get $15 for the round, if not you get $5.

In order to find a competing roll for *Task B* you will be matched with five other participants in the room. The five participants you can match with in *Task B* are randomly chosen at the start of the experiment, and these five participants will remain constant across all thirty rounds. However, because they are chosen randomly, and the matching is anonymous, you will never know which participants you interact with, nor will they know that they interacted with you.

The competing roll selected to face your roll in *Task B* is randomly selected from any of the matched five participants that also select *Task B* that round. However, in the event that **all five** matched participants chose *Task A* that round, the competing roll will instead be made by the computer, rolling a virtual version of the die at your desk.

Round Feedback. After you have made your task choice and rolled your die, you will be informed of the results for the round. If you chose *Task A*, the ball drawn from the urn will be reported to you, and you will be informed of whether you earned $15 or $5 for the round. If you chose *Task B*, you will be informed of the roll you competed against, and you will be informed of whether you earned $15 or $5 for the round. After this feedback the round will then end.

End of the Experiment. After you have completed 30 rounds, you will be asked to complete a brief survey. After completing the survey, two of the completed rounds in the experiment will be randomly selected for payment, where each round has the same chance of being selected. You will be paid based on your earnings from those two rounds, and only those two rounds.

There is no additional show-up payment for completing the study. However, given the $15 or $5 payment from each round, the minimum payment for completing the experiment is $10, while the maximum payment is $30.

Quiz. Before we start, you will be asked some questions to ensure you understand the experiment. Your answers to these questions will not directly affect your earnings, but simply serve to ensure you understand these instructions and how your decisions and die-rolls are used to determine your earnings.

B.2. **Instructions for Exogenous-None.**

Instructions for Today's Experiment

Introduction. Welcome to this experiment on decision-making. Please read these instructions carefully as they explain how you earn money from the decisions you make in this experiment. There will be no talking during today's session. If you have a question, please raise your hand and an experimenter will answer your question in private.

You will remain anonymous during the experiment and after it has concluded. Only a randomly assigned ID number will identify your decisions, and any research data collected during the course of the study will only identify your decisions by that number. The only personally identifiable information that will be recorded is that you participated in this session and were paid.

Your participation in this research study is completely voluntary. Should you change your mind about participating, you can withdraw from the study at any time. Your participation in the study will remain anonymous, and no identifying records will be retained of your withdrawal from the study. However, if you do decide to withdraw after or during the study, because of the anonymity inherent to the data collection, the researchers will be unable to delete the record of your responses.

Your current and future status with the University of Pittsburgh and any other benefits for which you qualify will be the same whether you participate in this study or not. If you withdraw during the experiment you are entitled to a $6 show-up payment, where the completed study payments (between $10 and $30) are only given to subjects who finish the study.

This study is being conducted by Alistair Wilson who can be reached at alistair@pitt.edu

Preliminaries.

- On the desk in front of you is a ten-sided die. Please put your hand up now if there is not a ten-sided die at your desk.
- We will explain in a moment how you will use this die in the experiment. For now, please roll the die several times to ensure that you are happy that it is a fair die and that you know how to read the outcome.
- The sides of the die have ten numbers ranging from 0 (the lowest) to 9 (the highest), where the numbers $\underline{6}$ (six) and $\underline{9}$ (nine) are each underlined, to avoid confusing them for one another.

Description of the Experiment. The experiment will consist of 30 rounds. In each round you need to do only one thing:

- Roll your die one time, and report the outcome.

Depending on the outcome of your die roll, you will receive a round payment of either $15 or $5.

Task: High Rolling. You will roll your die one time for the round and report the roll. After you report the result the computer then puts your roll in competition with *another* roll.

Your round earnings are calculated as follows:

- If your roll is higher than the competing roll you earn $15 for the round. For example if you rolled a 5 and the competing roll was between 0 and 4.
- If your roll is lower than the competing roll you earn $5 for the round. For example if you rolled a 5 and the competing roll was between 6 and 9.
- If the two rolls are the same, the computer randomly chooses one of the two rolls with equal probability (a 50-in-100 chance) to be the winner, as if flipping a fair coin. If your roll wins you get $15 for the round, if not you get $5.

The competing roll selected to face your roll is made by the computer, rolling a virtual version of the die at your desk.

Round Feedback. After you have rolled your die, you will be informed of the results for the round. Specifically, you will be informed of the roll you competed against, and you will be informed of whether you earned $15 or $5 for the round. After this feedback the round will then end.

End of the Experiment. After you have completed 30 rounds, you will be asked to complete a brief survey. After completing the survey, two of the completed rounds in the experiment will be randomly selected for payment, where each round has the same chance of being selected. You will be paid based on your earnings from those two rounds, and only those two rounds.

There is no additional show-up payment for completing the study. However, given the \$15 or \$5 payment from each round, the minimum payment for completing the experiment is \$10, while the maximum payment is \$30.

Quiz. Before we start, you will be asked some questions to ensure you understand the experiment. Your answers to these questions will not directly affect your earnings, but simply serve to ensure you understand these instructions and how your decisions and die-rolls are used to determine your earnings.