Lying Aversion and the Size of the Lie[*]

Uri Gneezy, Agne Kajackaite, and Joel Sobel[†]

Abstract

This paper studies lying. An agent randomly picks a number from a known distribution. She can then report any number and receive a monetary payoff based only on her report. The paper presents a model of lying costs that generates hypotheses regarding behavior. In an experiment, we find that the highest fraction of lies is from reporting the maximal outcome, but some participants do not make the maximal lie. More participants lie partially when the experimenter cannot observe their outcomes than when the experimenter can verify the observed outcome. Partial lying increases when the highest outcome is ex ante decreases.
Journal of Economic Literature Classification Numbers: D03; C90; C72. Keywords: lying; deception; experiments; behavioral economics.

[*]We thank Johannes Abeler, Martin Dufwenberg, Kiryl Khalmetski, Dominique Lauga, Isabel Marcin, Daniele Nosenzo, Collin Raymond, Mattia Saccoccio, Dirk Sliwka, Marcin Waligora, referees, the Co-Editor, and seminar participants at the California Institute of Technology, EUI, HKUST, Maastricht University, University of Amsterdam, University of Arizona, University of Cologne, UCSD, and Xiamen University for comments. This research was conducted with financial support from the German Science Foundation (DFG) through the research unit "Design & Behavior" (FOR 1371) and the Behavioral Economics Research Spearhead Grant (University of Amsterdam). Sobel thanks NSF for financial support.

[†]Rady School of Management, University of California, San Diego, La Jolla, CA 92093, U.S.A. and CREED, 1018 WB Amsterdam, NL. E-mail: ugneezy@ucsd.edu; Los Angeles Behavioral Economics Laboratory, University of Southern California, Los Angeles, CA 90098, U.S.A. Email: kajackai@usc.edu; Department of Economics, University of California, San Diego, La Jolla, CA 92093, U.S.A. E-mail: jsobel@ucsd.edu.

1 Introduction

Situations frequently arise in which people can lie about their private information. Although lying is common, compelling real-world and laboratory evidence shows that people sometimes avoid telling lies that would increase their material payoffs.[1]

Some honest behavior is an optimizing response to economic incentives; people are honest because dishonesty might lead to material punishments. Businesses may avoid making false claims because if caught they would face substantial penalties. Individuals in long-term relationships may resist opportunities to make short-term gains through lying in order to maintain profitable relationships. In these cases, honesty may be an optimal response for an agent who trades off the short-term benefits of lying with the long-term consequences. Although these situations are common and important, standard models suffice to describe them. Laboratory evidence suggests that in addition to these instrumental motivations, honesty has intrinsic value. This paper pursues the idea that there are intrinsic costs associated with lying and provides theory and evidence about the form of these costs.

Lies come in different sizes. Some lies are more plausible than others. Some lies have different implications on payoffs. The literature contains informal discussions of lying costs that identify different ways to measure the size of lies and its interaction with lying costs. For example, Mazar, Amir, and Ariely **?** and Fischbacher and Föllmi-Heusi **?** suggest that the marginal cost of a lie is increasing in the magnitude of a lie, leading to the prediction that individuals might lie a little bit, but not take full advantage of strategic opportunities. This paper introduces intrinsic costs of lying and systematically connects these costs to the size of the lie. It derives equilibrium behavior and generates testable predictions.

We argue that the intrinsic cost of lying depends on the size of the lie and identify three different ways to measure the size of the lie: the payoff dimension (monetary gains of lying), the outcome dimension (the distance between what the agent observes and what she says), and the likelihood

[1]Examples of experimental evidence include: Abeler, Becker, and Falk **?**, Abeler, Nosenzo, and Raymond **?**), Cohn, Fehr, and Maréchal **?**, Dreber and Johannesson **?**, Erat and Gneezy **?**, Evans, Hannan, Krishnan, and Moser **?**, Fischbacher and Föllmi-Heusi **?**, Gneezy **?**, Lundquist, Ellingsen, Gribbe, and Johannesson **?**, Hannan, Rankin, and Towry **?**, López-Pérez and Spiegelman **?**, Mazar, Amir, and Ariely **?**, Sutter **?**, and Shalvi, Dana, Handgraaf, and De Dreu **?**.

dimension (the ex ante probability that the agent's report is true). The following example illustrates the three dimensions of lying costs. Consider a situation in which the participant rolls an n-sided die and receives a positive payoff by reporting a five (and zero otherwise). One could vary the payoff dimension of lying costs by varying the payoff associated with reporting five. A dishonest report of five is a bigger lie in the payoff dimension if the reward for reporting five is bigger. To identify the outcome dimension of lying, one could ask how the frequency of reports of five varies with what the subject observes. If lying costs are associated with the outcome dimension, then one might conjecture that the closer their observation is to five, the more subjects lie; for example, they would be more likely to report five if they observed four than if they observed two.[2] Finally, to illustrate the likelihood dimension of lying costs, imagine changing the probability of rolling a five. The larger is n, the larger the lie on the likelihood dimension is to report five. That is, bigger lies are statements that are less likely to be true ex ante.

In Section **??**, we introduce a basic model that generates predictions regarding how sensitivity to the size of the lie on each of these three dimensions affects behavior. We assume that utility is the sum of three terms: the first is the monetary payoff, the second depends directly on the true state and the report (and indirectly on the monetary payoffs associated with these reports), and the third term depends on the probability that an observer believes the report is honest. The second term captures the outcome and payoff dimensions, whereas the third term captures the likelihood dimension.

Without the likelihood dimension, the theoretical model is a straightforward decision problem. Adding the likelihood dimension complicates the analysis because it adds a strategic aspect.[3] To incorporate the likelihood

[2]The experiment by Lundquist, Ellingsen, Gribbe, and Johannesson **?** is a good way to understand the outcome dimension. Their participants play a deception game (Gneezy **?**) in which first the Sender takes a test and then sends a message regarding the results of the test to the receiver. The Sender receives a fixed positive payoff if the receiver believes that she passed a certain threshold in her test. Because only two payoffs are possible (zero if not passing, and a fixed payment if passing), the size of the lie on the payoff dimension is constant. However, the size of the lie can be determined based on how close the sender's performance was to the actual threshold, which is how Lundquist et al. define the size of the lie: "We test whether the aversion to lying depends on the size of the lie (i.e. that the aversion to lying is stronger the further you deviate from the truth) . . . "

[3]Four recent papers, Abeler, Nosenzo, and Raymond **?**, Dufwenberg and Dufwenberg **?**, Garbarino, Slonim, and Villeval **?**, and Khalmetski and Sliwka **?**, introduce models that capture the likelihood dimension. We discuss these models after we present our formal

dimension, we follow Akerlof and Kranton **?**, Benabou and Tirole **?**, Tajfel **?**, Tajfel **?**, Tajfel and Turner **?**, and Turner and Onorato **?**, who argue that agents place an intrinsic value on “social identity.” These theories posit that the way others – even strangers – perceive an individual determines that individual’s social identity. Social identity concerns may influence an agent’s behavior even if she does not anticipate further interactions. In that sense, social identity is fundamentally different from traditional discussions of reputation. At the same time, these theories permit identities to be based on an internal notion of what is appropriate behavior. We assume that agents wish to be perceived as being honest and therefore gain utility from appearing honest. This utility could be instrumental (if people who are perceived to be honest get treated better). We focus, however, not on the instrumental utility of appearing to be honest but on the interpretation that being viewed as honest is an intrinsically valued part of an agent’s social identity. This intrinsic preference provides a motivation for an agent to sacrifice monetary payoffs in order to appear honest.

In Section **??**, we present the theoretical results of the model. The model makes a unique prediction. The equilibrium involves a cutoff value – if the agent draws an outcome above the cutoff, she never lies. If she draws an outcome below the cutoff, she may lie and, if she lies, she makes a claim above the cutoff. Furthermore, we find that higher claims are perceived to be less likely to be honest. Therefore, agents tell partial lies in equilibrium if social identity concerns are large enough. However, in equilibrium, dishonest claims of the maximal value arise with positive probability. Our most novel findings are qualitative results that capture the intuition that reducing the ex ante probability of the maximal outcome increases the frequency of partial lies. We interpret this finding as evidence that the likelihood dimension is an important part of the cost of lying. In Section **??**, we summarize the theoretical findings in a way that motivates the experimental portion of the paper.

In Section **??**, we describe the experimental design. Our experiments manipulate three characteristics of the game. First, we compare a game in which the experimenter may observe the subjects’ outcomes to one in which the experimenter cannot observe the outcome, even ex post. Second, to understand the outcome dimension of costs we vary the way in which the outcomes are labeled. Finally, we vary the prior distribution of outcomes to

results in Section **??**.

understand how the likelihood of outcomes affects behavior through social identity concerns.

Section ?? describes the experimental results. We find that people lie and a large fraction of those who do lie report the maximum lie. In addition, people report more partial lies when no one observes their outcomes than when outcomes can be observed. This finding is consistent with social identity concerns, because every lie leads to the worst possible social identity in the observed game, whereas the social identity is typically decreasing in the size of the lie in the game in which outcomes cannot be observed.[4] Another prediction that is consistent with the existence of social identity concerns is that participants tell more partial lies when the ex ante probability of the highest state decreases. Finally, with respect to the outcome dimension, we show that the fraction of dishonest reports does not depend on how outcomes are labeled, but the labels influence the frequency of partial lies.

2 Model

An agent's type consists of a pair (i, t), where $i = 1, \dots, N$ and $t \in [0, T]$, where N is a positive integer and $T > 0$. The value i represents the agent's observation, which can be thought of as the outcome of a roll of a die; the value t is the fixed cost of lying. The quantities i and t are independently distributed. $p_i > 0$ is the probability that the agent receives i. $F(\cdot)$ is the cumulative distribution function of t. The agent makes a claim, k, which is also assumed to be a number between 1 and N. We call an agent honest if she reports i when her type is of the form (i, t); she is dishonest (and her report is a lie) otherwise.[5] The agent receives a monetary reward of v_j if she reports j; if $i < j$, then $v_i < v_j$.

Denote the probability that an agent reports j given (i, t) by $s(j \mid i, t)$.[6] That is, an agent's strategy s maps type (i, t) into a probability distribution over reports.

[4]Stated differently, in the observed game, social identity is binary: partial lies and maximal lies lead to the same social identity. In the non-observed game, the social identity associated with telling the maximal lie is strictly less than that of a partial lie.

[5]In this model, the agent's claim is about her observation. Lying about intentions is not possible in our framework. The question of whether the lying-cost structure is the same in such different cases is left for future research.

[6]$s(j \mid i, t) \geq 0$ and $\sum_{j=1}^{N} s(j \mid i, t) = 1$ for all i and t.

We assume that an agent's preferences depend on three elements: the monetary payoff, costs associated with the relationship between her value i and her report j, and the extent to which her behavior influences her social identity. We assume that the utility function takes the form

$$v_j - C(i,j,t) + \beta\gamma_{ij}(s) \tag{1}$$

for $\beta > 0$, but we limit the analysis to a special case of this functional form by imposing restrictions on the second and third terms.

First, we assume

$$C(i,j,t) = \begin{cases} 0 & \text{if } i = j \\ t + c(i,j) & \text{if } i \neq j \end{cases}.$$

$C(\cdot)$ represents the direct cost of lying. We measure the social identity cost with the third term. We assume that the agent values honest behavior or being perceived as behaving honestly. The perception can be part of the agent's self image or it could be part of how the agent is viewed by others (social identity). The function $\gamma_{ij}(\cdot)$ captures this element of preferences. We specialize the expression in (**??**) by assuming γ_{ij} takes the form:

$$\gamma_{ij}(s) = \lambda(I_{ij} - 1) + (1-\lambda)\rho_j(s), \tag{2}$$

where

$$I_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases},$$

$\lambda \in [0,1]$, and $\rho_j(s)$ is the probability that a report of j is interpreted as being honest. $\rho_j(s)$ depends on the strategy profile s and is computed using Bayes's Rule:

$$\rho_j(s) = \frac{h_j(s)}{h_j(s) + r_j(s)}, \tag{3}$$

where

$$h_j(s) = \int_0^T s(j \mid j,t)dF(t)p_j \tag{4}$$

is the probability that j is reported honestly and

$$r_j(s) = \sum_{i=1, i\neq j}^{N} \int_0^T s(j \mid i,t)dF(t)p_i \tag{5}$$

is the probability that j is reported dishonestly. It must be that $r_j(s) \geq 0$ for all j. If $h_j(s) + r_j(s) = 0$, then $\rho_j(s)$ is not defined. This possibility does not arise because our assumptions on the distribution of lying costs will guarantee $h_j(s) > 0$.

The definition of $\gamma_{ij}(\cdot)$ in (**??**) spans two situations. When $\lambda = 1$, the social identity term depends only on whether $i = j$. This specification is relevant when the agent's value i is publicly observed and no uncertainty exists regarding whether a report is dishonest, so it models the "observed game" experiments we conduct. When $\lambda = 0$, the social identity term depends only on the inferences an observer can make knowing the reported value j and the strategy s, without knowing the observation i. This specification applies to the "non-observed game" experiments, provided the agent's social identity depends only on how others perceive her behavior. When $\lambda \in (0, 1)$, our specification includes both a self-assessment of social identity and the perception of others. Because we view both self-assessment and perception as potentially relevant, these cases are the focus of our analysis. An agent with correct beliefs about whether she is observed and whose social identity depends only on how others perceive her sets $\lambda = 0$ in the non-observed game and $\lambda = 1$ in the observed game. Even if the agent cares only about how others view her behavior, values of λ strictly between zero and one may be appropriate in the non-observed game. For example, the agent may have incorrect beliefs about the beliefs of others (and, in particular, may believe that the observer knows what she knows).

Given our specifications of $C(\cdot)$ and $\gamma_{ij}(\cdot)$, we can rewrite the utility of a type (i, t) agent who reports j as:

$$U(i, j, t; s) = \begin{cases} v_j + \beta(1 - \lambda)\rho_j(s) & \text{if } i = j, \\ v_j - c(i, j) - t - \beta\lambda + \beta(1 - \lambda)\rho_j(s) & \text{if } i \neq j. \end{cases} \tag{6}$$

We also maintain additional assumptions on the cost function $c(\cdot)$ throughout our analysis. We assume that $c(\cdot)$ is nonnegative, $c(i, i) = 0$, weakly increasing in $\mid i - j \mid$ and $c(i, j) + c(j, k) \geq c(i, k)$. We also assume that $c(N - 1, N) < v_N - v_{N-1}$. These conditions include as a special case a model in which there is a categorical cost of lying ($C(i, j) = 0$ if $i = j$ and otherwise the lying cost is positive and independent of i and j). We use condition $c(N - 1, N) < v_N - v_{N-1}$ to avoid the uninteresting case in which the cost of lying is so great that no one wishes to make the highest report dishonestly.

With these assumptions, we can normalize the lying costs and define α as $\beta(1-\lambda)$. These considerations lead to the functional form for preferences that we use for the analysis:

$$U(i,j,t;s) = \begin{cases} v_j + \alpha\rho_j(s) & \text{if } i = j, \\ v_j - c(i,j) - t + \alpha\rho_j(s) & \text{if } i \neq j. \end{cases} \tag{7}$$

(The "t" specification in (**??**) replaces $t+\beta\lambda$ in (**??**); this is the normalization of fixed costs of lying.)

Finally, we assume that the distribution over types $F(\cdot)$ is continuous, has support $[0,T]$, $F(0)=0$, and that $T > v_i - v_{i-1}$ for all $i = 1,\ldots,N$.[7] Combined, these conditions guarantee that almost all agents find lying costly and that some agents find lying so costly that they will never lie. In particular, they guarantee the denominator in the expression for $\rho_j(\cdot)$, (**??**), is always positive.

Some properties depend on whether the parameter α in (**??**) is zero or strictly positive. The case $\alpha = 0$ corresponds to when the observer knows the agent's true value. Increases in α correspond to the agent placing increasing weight on appearing honest to individuals who do not observe the agent's true value.

Representation (**??**) is special. It leaves out factors that may be important.[8] Distributional concerns may play a role, but no other active agents are in our experiments. Our specification assumes that the subject does not care about the experimenter's monetary payoff. The social identity term is a reduced form that captures some concerns the subject has about how she is perceived. Additive separability, homogeneous preferences over monetary payments, and risk neutrality may be important restrictions.[9] The assumption that $c(i,j)+c(j,k) \geq c(i,k)$ simplifies our analysis because it guarantees

[7]Our results do not require that 0 is in the support of $F(\cdot)$, but it simplifies the exposition to be able to guarantee that a type $(N-1,t)$ agent will lie if t is sufficiently small. This property holds for t near zero if $c(N-1,N) < v_N - v_{N-1}$. If $[\underline{t},T]$ is the support of $F(\cdot)$ for $\underline{t} > 0$, then we can assume instead that $c(N-1,N)+\underline{t} < v_N - v_{N-1}$. We do make use of the property that $F(0)=0$, which guarantees that $t > 0$ with probability one.

[8]For example, Charness and Dufwenberg **?** study how promises and the desire to keep promises influence honest behavior. Erat and Gneezy **?** study the decision to behave honestly when distributional concerns are present. Kajackaite **?** shows that the intrinsic cost of lying about ability is higher than the cost of lying about luck. Marcin **?** argues that agents may report honestly to signal ability.

[9]One can view v_k as measured in utils, so risk neutrality is not restrictive given sepa-

that if anyone dishonestly reports k, then no one who observes k will be dishonest.[10] Assuming t enters the cost function separably means, conditional on wanting to lie, the preferences of type (i, t) do not depend on t. The notation suppresses the possible dependence of costs on the monetary payoffs v_i, we imagine that if v_j changes, the cost of claiming j dishonestly would change. The cost function $c(i, j)$ is defined for $i > j$, but we will show that in equilibrium no agent will report less than what she observes. We chose this specification because it is tractable – in particular, it permits us to show (in Lemma **??**) that ρ_i is uniquely determined in equilibrium when $\alpha > 0$ – and rich enough to provide testable hypotheses.

We analyze an equilibrium in this setting, which consists of strategies $s(k \mid i, t)$ such that

(a) $s(k \mid i, t) \geq 0$ for all k, i, t and $\sum_{j=1}^{N} s(j \mid i, t) = 1$ for all i and t,

(b) $s(k \mid i, t) > 0$ only if k maximizes (**??**) (with respect to j),

(c) $\rho_j(s)$ is computed using (**??**) and (**??**).

There exists t^* large enough such that all types of the form (i, t) with $t > t^*$ would prefer to report i instead of $j \neq i$ independent of ρ. We assume that $t^* < T$, which guarantees that the set of t such that $s(i \mid i, t) > 0$ has positive probability for all i. Consequently, the denominator in (**??**) is strictly positive.

The existence of an equilibrium in which each type (i, t) plays a pure strategy follows from standard arguments (Schmeidler **?**).[11]

The game we study posits that there exists a single agent who makes an observation and decides what to report by maximizing utility with respect to beliefs. It is formally equivalent to view the game as one in which different agents make observations. This point of view is consistent with our experimental design, but it is arguably less plausible to assume that ρ_j is independent of (i, t) when (i, t) represents the type of a single subject rather than the characteristics of a member of a large population. At the

rability and homogeneity.

[10]We find that our data are consistent with the prediction that if anyone dishonestly reports k, then no one who observes k will be dishonest.

[11]Schmeidler's theorem guarantees the existence of a pure-strategy equilibrium in a game in which a continuum of players exists and each player has a finite action set. The result applies to our game by treating each type as a player.

same time, we have no evidence or theoretical reason to think the beliefs of different individuals should be systematically different. For this reason, we maintain the (equilibrium) assumption that the interpretation of reports does not vary with (i, t) and view our model as an useful representation of the experimental game.

The definition of equilibrium requires that beliefs are statistically correct (property (c)). This assumption is less important for our conclusions. We have pointed out that the subject can have incorrect beliefs about the probability that she is observed (formally, she can have incorrect beliefs about λ).

3 Analysis

This section describes properties of equilibria. The first main property is that the social identities generated in equilibrium must be unique. We also show how equilibrium behavior depends on the prior distribution.

We denote strategies by s (or s', s''), the associated social identities by ρ_k (or ρ'_k, ρ''_k), and the utility without lying costs of a report by W_k (or W'_k, W''_k) so that $W_k = v_k + \alpha\rho_k$. The first result identifies a structural property of equilibrium strategies:

Proposition 1 *If there exists $t' > 0$ such that $s(k \mid j, t') > 0$ for $k \neq j$, then for all $i \neq j$ and $t > 0$, $s(j \mid i, t) = 0$.*

In words, Proposition **??** states that if some agent finds lying valuable when the true state is j, then no agent will dishonestly report j.

Proof. If $s(k \mid j, t') > 0$, then

$$W_k - C(j, k, t') \geq W_j. \tag{8}$$

To prove the lemma, it suffices to show that for all t,

$$W_k - C(i, k, t) > W_j - C(i, j, t). \tag{9}$$

By inequality (**??**), $W_k - W_j \geq C(j, k, t')$. However,

$$C(j, k, t') > c(j, k) \geq c(i, k) - c(i, j) = C(i, k, t) - C(i, j, t), \tag{10}$$

where the strict inequality follows from the definition of $C(\cdot)$ when $t' > 0$ and the equation follows (for all t) by the definition of $C(\cdot)$. The weak inequality holds because $c(i,j) + c(j,k) \geq c(i,k)$. Inequality (**??**) follows immediately from inequalities (**??**) and (**??**). ■

Proposition **??** implies no observed outcome j exists with the property that some type (j, t') would lie while another would dishonestly report j. To get an intuition for the result, consider the leading case where $k > j > i$. The maintained assumptions on $C(\cdot)$ imply that $C(j,k,t') > C(j,k,t) - C(i,j,t)$. Two features of the cost function lead to the proposition. First, if a type is willing to pay the fixed cost (to report k instead of j), then no other type that has already decided to lie would not find it optimal to report j. The second feature of the cost function is that $c(i,j) + c(j,k) \geq c(i,k)$ so that marginal cost of increasing the size of a lie (in outcome space) is non-increasing.

Proposition **??** has two consequences. The first consequence (Corollary **??**) is that if there exists a type (j,t) that lies, then observers (who do not know the observed value) think that anyone who reports j is honest. The first consequence follows directly from the proposition because the proposition states that if there exists a type (j,t) who lies, then no one dishonestly claims j. Hence, any report of j must be honest. The second consequence is that no one ever makes a claim that is less than the truth. The second consequence (Proposition **??**) follows because if $j > i$, then $v_j > v_i$. If type j reports i, then $\rho_j = 1$. So $W_j > W_i$ and type (j,t) would be better off reporting honestly than reporting i.

Corollary 1 *If $s(k \mid k,t) < 1$ for some t, then $\rho_k = 1$.*

Proof. Proposition **??** implies that if $s(k \mid k,t) < 1$, then the probability that another type reports k, r_k, is equal to zero. The result follows from the definition of ρ_k (given in equation (**??**)) and $h_k \neq 0$. ■

Proposition 2 *If $i < j$, then $s(i \mid j,t) = 0$ for all t.*

Proof. If $s(i \mid j,t) > 0$ for $j \neq i$, then $W_j \leq W_i - C(j,i,t)$ and, by Corollary **??**, $\rho_j = 1$. It follows that $v_i > v_j$, and therefore $i > j$. ■

The model provides a unique equilibrium prediction in the sense that all equilibria give rise to the same set of values for ρ_k. The next result implies that the set of claims that are made dishonestly does not depend on the equilibrium selected.

Lemma 1 *Suppose $\alpha > 0$. If s' and s'' are two equilibria, then $\rho'_k = \rho''_k$ for all k.*

The lemma requires that social identity matters ($\alpha > 0$). If $\alpha = 0$, then $\rho_1, \ldots, \rho_N$ do not influence preferences, so uniqueness of these values is not important. Equilibrium utilities are unique for all α.[12]

An intuition for the result follows. Suppose two equilibria exist that give rise to different ρ. Suppose that moving from the first equilibrium to the second, the social identity for reporting k goes down by the most over all possible reports. Thus, reporting k dishonestly in the second equilibrium is less attractive. However, if the probability of dishonest reports of k in the second equilibrium is lower, then the value associated with reports of k must be higher in the second equilibrium. (This observation requires that the number of honest reports of k does not go down, which follows from Proposition **??**.) Consequently, all ρ_k must be higher in the second equilibrium. One can use the same argument to show that all ρ_k are higher in the first equilibrium. Consequently, the proposition must hold.

Proof. Let M be the set of minimizers of $\rho'_j - \rho''_j$. If

$$W'_k - C(i,k,t) \geq W'_j - C(i,j,t),$$

then $W''_k - C(i,k,t) + \alpha\left((\rho''_j - \rho'_j) - (\rho''_k - \rho'_k)\right) \geq W''_j - C(i,j,t)$. Hence, if $k \in M$, then $W''_k - C(i,k,t) \geq W''_j - C(i,j,t)$ with strict inequality unless $j \in M$. It follows that if $s'(k \mid i,t) > 0$ for $i \neq k$, $k \in M$, then $s''(j \mid i,t) = 0$ for all $j \notin M, j \neq i$. Hence, $\sum_{k \in M} r'_k \leq \sum_{k \in M} r''_k$ and for at least one $k \in M, r'_k \leq r''_k$. By Proposition **??**, it follows that $\rho'_k \geq \rho''_k$, and hence $\rho'_i \geq \rho''_i$ for all i. Because we can use the same argument reversing the roles of the two equilibria, it follows that $\rho'_i = \rho''_i$ for all i. ■

Equilibrium may not be unique. Consider the special case in which $c(i,j) = 0$. In this case, all agents have identical preferences over lies. If equilibrium involves $\rho_k < 1$ for more than one value of k, then there will

[12] Even when $\alpha = 0$ the ρ_j are uniquely determined in equilibrium for "most" parameter values. Formally, treat the parameter values, $c(i,j)$, $i < j$, $i = 1, \ldots, N-1$, $j = 2, \ldots, N$ and v_i $i = 1, \ldots, N$ as elements of $\mathbb{R}^{N(N-1)/2}$. A type (i,t) agent either tells the truth or selects the report $j \neq i$ that maximizes $v_j - c(i,j)$. There exist distinct i,j,k such that $v_j - c(i,j) = v_k - c(i,k)$ for a subset of parameter values that is closed and has Lebesgue measure zero. Consequently, equilibrium behavior (and therefore equilibrium values of ρ_i) is generically unique even when $\alpha = 0$.

typically be different signaling strategies compatible with equilibrium. Nevertheless, Lemma **??** guarantees that the conditional probability of a lie given the observed value and the conditional probability of a report being honest are uniquely determined in equilibrium.

Let s be an equilibrium and let

$$L(s) = \{k : \text{ there exists } i \neq k \text{ and } t, \text{ such that } s(k \mid i, t) > 0\}.$$

$L(\cdot)$ is the set of claims that are made dishonestly with positive probability in equilibrium. Note that $\rho_k(s) < 1$ if $k \in L(s)$ and, by Proposition **??**, $\rho_k(s) = 1$ if $k \notin L(s)$, so $L(s) = \{k : \rho_k(s) < 1\}$.

Proposition 3 *The highest claim is made dishonestly with positive probability.*

Proof. Recall that N is the highest value. If $N \notin L(s)$, then $\rho_N(s) = 1$ and $W_N - W_{N-1} \geq v_N - v_{N-1}$. Because $c(N-1, N) < v_N - v_{N-1}$ by assumption, for t sufficiently small $W_N - C(N-1, N, t) > W_{N-1}$, so a $(N-1, t)$ agent would prefer to dishonestly report N than to tell the truth. Because Corollary **??** implies that a dishonest agent will never underreport, the result follows. ■

Proposition **??** demonstrates that maximal lies (reporting N) occur with positive probability. When social identity matters, partial lies will occur in equilibrium. That is, N will not be the only claim reported dishonestly under intuitive conditions.

Proposition 4 *Suppose $N > 2$ and $v_{N-1} - v_{N-2} > c(N-2, N-1)$. If either*

1. *$\alpha > v_N - v_{N-1}$ and p_N is sufficiently small or*

2. *α is sufficiently high,*

then N is not the only claim made dishonestly with positive probability.

Proof. We show $L(s) = \{N\}$ is not possible if the conditions in the proposition hold. If $L(s) = \{N\}$, then $\rho_k(s) = 1$ for $k < N$ and $(N-2, t)$ must report either $N-2$ or N. Consequently, for all t,

$$\max\{v_{N-2} + \alpha, v_N + \alpha\rho_N - C(N-2, N, t)\} \geq v_{N-1} + \alpha - C(N-2, N-1, t). \tag{11}$$

Because $v_{N-1} - v_{N-2} > c(N-2, N-1)$, there exists $\tilde{t} > 0$ such that if $t < \tilde{t}$,

$$v_{N-1} + \alpha - C(N-2, N-1, t) > v_{N-2} + \alpha. \tag{12}$$

It follows from (**??**) that

$$v_N + \alpha\rho_N - C(N-2, N, t) \geq v_{N-1} + \alpha - C(N-2, N-1, t) \tag{13}$$

and that the probability that $N-2$ reports N is at least $F(\tilde{t}) > 0$ for all α and p_N. Consequently, there exists $b < 1$ such that $\rho_N < b$. We now have a contradiction: when Condition (**??**) holds, ρ_N must converge to 0 and (**??**) contradicts $\alpha > v_N - v_{N-1}$; when Condition (**??**) holds, (**??**) cannot hold if α approaches infinity. ■

Propositions **??** and **??** require conditions on $v_k - v_{k-1}$. Without assumptions on the rate of increase of the rewards, some claims may not be worth lying for. For the next result, we impose a stronger condition: $v_k - v_{k-1} > c(k-1, k)$ for all k. In many experimental designs, $v_k - v_{k-1}$ is a positive constant. This condition holds in the models of Dufwenberg and Dufwenberg **?** and Khalmetski and Sliwka **?**. Clearly, if $v_k > v_{k-1}$ and $c(k-1, k) = 0$, then $v_k - v_{k-1} > c(k-1, k)$. This assumption adds structure to the equilibrium.

Proposition 5 *Suppose $v_k - v_{k-1} > c(k-1, k)$ for all k. There exists $n^* < N$ such that $L(s) = \{k : k > n^*\}$.*

$L(s) = \{k : k > n^*\}$ is equivalent to $\rho_k < 1$ for $k > n^*$ and $\rho_k = 1$ for $k \leq n^*$.

Proof. Let $n^* = \max\{k : \rho_k = 1\}$. It follows from Proposition **??** that $\rho_1 = 1$, so n^* is well defined. We must show $\rho_k = 1$ for $k \leq n^*$. We know that $\rho_{n^*} = 1$. Assume that $\rho_k = 1$ for $k = k^*, \ldots, n^*$ and $k^* > 1$. The condition

$$v_j - v_{j-1} > c(j-1, j) \text{ for all } j \tag{14}$$

implies that for t sufficiently small, $(k^* - 1, t)$ strictly prefers to report k^* to $k^* - 1$. Hence, by Corollary **??**, $\rho_{k^*-1} = 1$, which establishes the result. ■

When $\alpha = 0$, the conclusion of Proposition **??** is stronger. Condition (**??**) and the maintained assumption that $c(i, k-1) + c(k-1, j) \geq c(i, k)$ implies that $v_k - v_{k-1} > c(i,k) - c(i,k-1)$. Consequently, (**??**) implies that $v_k - C(i,k,t) > v_{k-1} - C(i,k-1,t)$. Hence, in the observed game, any agent who lies reports N and so no partial lies occur in the observed game ($n^* = N-1$) when (**??**) holds.

Proposition **??** states that a cutoff observation exists. If the outcome is above this cutoff, then agents never lie. If the outcome is below the cutoff, then they lie with positive probability and dishonest claims are above the cutoff.

If agents care about their social identity, then the prior distribution over outcomes should influence behavior in a systematic way. In particular, if the prior distribution shifts mass from the most likely profitable outcome N to lower outcomes, telling the biggest lie should become less attractive. The next result formalizes this intuition. We consider a simple shift of probabilities: The distribution $p'' = (p''_1, \ldots, p''_N)$ is a **proportional shift from N** of $p' = (p'_1, \ldots, p'_N)$ if there is $\lambda \in (0,1)$ such that $p''_N = \lambda p'_N$ and $p''_i = (1 - \lambda p'_N)p'_i/(1 - p'_N)$ for $i < N$.

The next result is our key comparative-statics result. Compare a situation in which the observed outcomes are ex ante equally likely to one in which the observation giving the highest reward is extremely unlikely. In the second case, if all dishonest agents make the highest claim, then lower social identity would result. If being viewed as honest is sufficiently valuable, then dishonest agents would prefer to make a smaller claim, losing some monetary payment, but gaining a stronger social identity for being honest.

Proposition 6 *Suppose $v_k - v_{k-1} > c(k-1,k)$ for all k. Let p'' be a proportional shift from N of p'. Let s' (s'') be an equilibrium associated with a prior probability distribution p' (p''). For each i, $\rho'_i \geq \rho''_i$, $L(s') \subset L(s'')$, and for all $k < N$, the probability of a dishonest report of k is at least as great under p'' as under p'.*

Three conclusions follow from Proposition **??**. The first conclusion is that shifting prior probability to less valuable outcomes lowers the probability that any claim is viewed as honest. Hence, a proportional shift from N lowers the utility of the agent. It is intuitive that the shift should lower ρ_N. If p'' is a proportional shift from p', then $\rho'_N < \rho''_N$ suggests reporting N under p'' is more attractive than under p',which implies $\rho'_N \geq \rho''_N$. The second conclusion is that more claims are made dishonestly under s'' than under s'. Thus, shifting probability from N makes the subject willing to dishonestly report lower claims. Because a cutoff observation exists (by Proposition **??**), shifting probability from N lowers the lowest value that is reported dishonestly. Loosely, the social identity loss associated with higher claims could be large enough to convince the subject to make a more modest lie. The third claim is that the probability of partial lies (reports that are both dishonest and less than N) increases after a proportional shift. One reason for this change is non strategic. If p'' is a proportional shift of p', then there is more prior probability on low outcomes. Hence, there are more situations under which lying is attractive. The conclusion depends on more than this observation. The first conclusion implies that an agent's social identity will be lower after the proportional shift. The lower social identity decreases the incentive to lie (and acts against the third conclusion). Further, some of the lies are maximal lies. There might be more lies under p'', but not more partial lies.

One might conjecture that a proportional shift from N would lead to a reduction in the fraction of maximal lies. We cannot establish this property. Because the proportional shift creates more possible lies (since a larger fraction of outcomes is less than N) the ex ante fraction of maximal lies might increase even as the payoff of these lies decreases. Hence an asymmetry exists between claims of N and claims of $k < N$. The reason for this asymmetry is that if a smaller fraction of subjects dishonestly report $k < N$ under p'' than under p', then $\rho''_k > \rho'_k$, because $p''_k > p'_k$, whereas $\rho'_N \geq \rho''_N$ is possible even if a smaller fraction of subjects dishonestly report N under p'' than under p', because $p'_N > p''_N$.

Proof. Let l'_{ik} (l''_{ik}) be the conditional probability that an agent who observes i dishonestly reports k under p' (p''). Hence, $r'_k = \sum_{i=1}^{N-1} p'_i l'_{ik}$ and $r''_k = \sum_{i=1}^{N-1} p''_i l''_{ik} = (1 - \lambda p'_N) \sum_{i=1}^{N-1} p'_i l''_{ik}/(1 - p'_N)$ by the definition of p'' (note that because $l_{NN} = 0$, there is no term involving $p_N l_{NN}$).

If k is reported dishonestly, then no one who observes k lies, so $h_k = p_k$ by Proposition **??**. It follows that if k is claimed dishonestly under s', $\rho'_k \geq \rho''_k$ if $r'_k/p'_k \leq r''_k/p''_k$; and if k is claimed dishonestly under both s' and s'', then $\rho'_k \geq \rho''_k$ if and only if $r'_k/p'_k \leq r''_k/p''_k$

Hence, if $k \in L(s'), k < N$, then $\rho'_k \geq \rho''_k$ if $\sum_{i=1}^{N-1} p'_i l'_{ik} \leq \sum_{i=1}^{N-1} p'_i l''_{ik}$,

$$\text{if } k \in L(s') \cup L(s''), k < N, \text{ then } \rho'_k \geq \rho''_k \text{ if and only if } \sum_{i=1}^{N-1} p'_i l'_{ik} \leq \sum_{i=1}^{N-1} p'_i l''_{ik}, \tag{15}$$

and

$$\rho'_N \geq \rho''_N \text{ if and only if } \sum_{i=1}^{N-1} p'_i l'_{iN} \leq \frac{(1-\lambda p'_N)\sum_{i=1}^{N-1} p'_i l''_{iN}}{\lambda(1-p'_N)}. \tag{16}$$

Because $\lambda(1-p'_N) \leq 1-\lambda p'_N$, it follows from inequalities (**??**) and (**??**) that if $s'(k \mid i,t) > 0$ for some (i,t), then $\rho'_k \geq \rho''_k$ if

$$\sum_{i=1}^{N-1} p'_i l'_{ik} \leq \sum_{i=1}^{N-1} p'_i l''_{ik}. \tag{17}$$

Let M be the set of minimizers of $\rho'_j - \rho''_j$. If

$$W'_k - C(i,k,t) \geq W'_j - C(i,j,t),$$

then $W''_k - C(i,k,t) + \alpha\left((\rho''_j - \rho'_j) - (\rho''_k - \rho'_k)\right) \geq W''_j - C(i,j,t)$. Hence if $k \in M$, then $W''_k - C(i,k,t) \geq W''_j - C(i,j,t)$ with strict inequality unless $j \in M$. Consequently, if $s'(k \mid i,t) > 0$ for $i \neq k$, $k \in M$, then $s''(j \mid i,t) = 0$ for all $j \notin M, j \neq i$.

Hence, $\sum_{k\in M} l'_{ik} \leq \sum_{k\in M} l''_{ik}$ and therefore

$$\sum_{i=1}^{N-1}\sum_{k\in M} p'_i l'_{ik} \leq \sum_{i=1}^{N-1}\sum_{k\in M} p'_i l''_{ik}.$$

It follows that for at least one $k \in M$,

$$\sum_{i=1}^{N-1} p'_i l'_{ik} \leq \sum_{i=1}^{N-1} p'_i l''_{ik}. \tag{18}$$

If follows from inequalities (**??**) and (**??**) that $\rho'_k \geq \rho''_k$ for some $k \in M$. By the definition of M, it must be that $\rho'_j \geq \rho''_j$ for all j. This establishes the first part of the proposition. The inclusion $L(s') \subset L(s'')$ follows from the definition of L and Proposition **??**. Because $p'_k \leq p''_k$ for $k < N$, inequality (**??**) implies that there are at least as many partial lies to k, under p'' as under p' for all k. ■

Proposition **??** describes what happens if one shifts probability from the most attractive observation. It demonstrates that such a shift increases the fraction of partial lies. Is this qualitative feature a consequence of the highest state having low absolute probability (p_N small) or low relative probability (p_i/p_N large for $i < N$)? Proposition **??** suggests that reductions in the absolute probability of the most attractive outcome lead to increases in the fraction of partial lies. The next proposition confirms this observation. For the proposition, we compare two environments. In one, the outcome is uniformly distributed over $1, 2, \ldots, N$. In the second, we permit observations that are not integers and, specifically, that the outcome is uniformly distributed over $.5, 1, \ldots, N - .5, N$. Hence, one moves from the first environment to the second by doubling the number of possible observed outcomes.[13]

Proposition 7 *Suppose $v_k - v_{k-.5} > c(k - .5, k)$ for all $k = 1, 1.5, \ldots, N$. Let p' be a uniform distribution on $\{1, 2, \ldots, N\}$ and Let p'' be a uniform distribution on $\{.5, 1, \ldots, N\}$. Let s' (s'') be an equilibrium associated with a prior probability distribution p' (p''). For each $i = 1, \ldots, N$, $\rho'_i \geq \rho''_i$ and $L(s') \subset L(s'')$. The probability of a dishonest report of $k < N$ is at least as great under p'' as under p'.*

The conclusion that $L(s') \subset L(s'')$ means that splitting observed outcomes lowers the lowest value that is reported dishonestly. For example, if observations 8, 9, and 10 are reported dishonestly when $N = 10$, then when observed outcomes are split, we would expect to see dishonest reports of 7.5, 8, 8.5, 9, 9.5, and 10.

If the only dishonest claim made is $\{N\}$ (that is, $L(s'') = \{N\}$), then more agents might report this claim dishonestly under p'' because the probability of an observation less than N is greater under p'' than under p'. This possibility could happen if α is so low that nearly all agents make the maximum lie. In general, we cannot rule out the possibility that maximal lies are more common under p'' for the same reason we could not do so in Proposition **??**: Under p'', outcomes less than N are more common and hence there is a higher ex ante probability of lying (and, perhaps, maximal lying).

Proposition **??** states that splitting observed outcomes lowers the threshold below which subjects lie. That is, $\rho'_i \leq \rho''_{i-1}$ and if $L(s') = \{k : k \geq n^*\}$ then $L(s'') \subset \{k : k \geq n^* - 1\}$. A careful examination of the argument demonstrates that when observed outcomes are split, reports need not change by more than one unit. That is, if (i, t) reports k prior to the split, then (i, t) would report k or $k - .5$ after the split.

[13]The proposition uses a specific notion of splitting that is consistent with our experimental design.

Proposition **??** requires a preliminary result, which gives conditions under which social identity is weakly decreasing.

Lemma 2 *Suppose $v_j - v_{j-1} > c(j-1, j)$ for all j. Social identity is weakly decreasing in report. That is, if $k' > k$, then $\rho_{k'} \leq \rho_k$.*

Proof. Proposition **??** implies that there exists n^* such that $\rho_j = 1$ if and only if $j \leq n^*$. We claim that if $j \geq n^*$, then $\rho_{j+1} < \rho_j$. This claim is sufficient to establish the proposition. The claim is true when $j = n^*$. When $N > j > n^*$, there must be a type (i, t) that dishonestly reports j. Hence, j solves $\max_k W_k - C(i, k, t)$ and, in particular,

$$W_j - C(i, j, t) \geq W_{j+1} - C(i, j+1, t), \tag{19}$$

which implies

$$\alpha(\rho_j - \rho_{j+1}) \geq v_{j+1} - v_j + c(i, j) - c(i, j+1) > c(j, j+1) + c(i, j) - c(i, j+1) \geq 0,$$

where the first inequality follows from (**??**), the second inequality by $v_{j+1} - v_j > c(j, j+1)$, and the third by the maintained assumption on $c(\cdot)$. It follows that $\rho_j > \rho_{j+1}$, which establishes the result. ■

We commented earlier that $v_j - v_{j-1} > c(j-1, j)$ for all j implies no partial lies arise when $\alpha = 0$. A stronger version of Lemma **??** holds in this situation: $\rho_k = 1$ for $k < N$ and $\rho_N < 1$.

Proof of Proposition ??. Let l'_{ik} (l''_{ik}) be the conditional probability that an agent who observes i dishonestly reports k under p' (p''). Hence, $r'_k = \sum_{i=1}^{N-1} l'_{ik}/N$ and $r''_k = \sum_{i=1}^{N-1} l''_{ik}/2N$.

If k is reported dishonestly, then no one who observes k lies, so $h_k = p_k$ by Proposition **??**. It follows that if k is claimed dishonestly under s', $\rho'_k \geq \rho''_k$ if $r'_k \leq 2r''_k$ and if k is claimed dishonestly under both s' and s'' then $\rho'_k \geq \rho''_k$ if and only if $r'_k \leq 2r''_k$. Hence, if $k \in L(s')$, then $\rho'_k \geq \rho''_k$ if $\sum_{i=1}^{N-1} l'_{ik} \leq \sum_{i=1}^{N-1} l''_{ik}$,

$$\text{if } k \in L(s') \cup L(s''), \text{ then } \rho'_k \geq \rho''_k \text{ if and only if } \sum_{i=1}^{N-1} l'_{ik} \leq \sum_{i=1}^{N-1} l''_{ik}. \tag{20}$$

It follows from inequality (**??**) that if $s'(k \mid i,t) > 0$ for some (i,t), then $\rho'_k \geq \rho''_k$ if

$$\sum_{i=1}^{N-1} l'_{ik} \leq \sum_{i=1}^{N-1} l''_{ik}. \tag{21}$$

Let M be the set of minimizers of $\rho'_j - \rho''_j$. If

$$W'_k - C(i,k,t) \geq W'_j - C(i,j,t),$$

then $W''_k - C(i,k,t) + \alpha\left((\rho''_j - \rho'_j) - (\rho''_k - \rho'_k)\right) \geq W''_j - C(i,j,t)$. Hence if $k \in M$, then $W''_k - C(i,k,t) \geq W''_j - C(i,j,t)$ with strict inequality unless $j \in M$. It follows that if $s'(k \mid i,t) > 0$ for $i \neq k$, $k \in M$, then $s''(j \mid i,t) = 0$ for all $j \notin M, j \neq i$. Hence, $\sum_{k \in M} l'_{ik} \leq \sum_{k \in M} l''_{ik}$ and therefore

$$\sum_{i=1}^{N} \sum_{k \in M} l'_{ik} \leq \sum_{i=1}^{N} \sum_{k \in M} p_i l''_{ik} \tag{22}$$

and for at least one $k \in M$,

$$\sum_{i=1}^{N-1} l'_{ik} \leq \sum_{i=1}^{N-1} l''_{ik}. \tag{23}$$

It follows from inequalities (**??**) and (**??**) that $\rho'_k \geq \rho''_k$ for some $k \in M$. By the definition of M, it must be that $\rho'_j \geq \rho''_j$ for all $j = 1, 2, \ldots, N$. Consequently, inequality (**??**) implies inequality (**??**) holds for all k. This establishes the first part of the proposition. The inclusion $L(s') \subset L(s'')$ follows from the definition of L and Proposition **??**. It remains to show that for $i < N$, $\sum_{i=1}^{N-1} l'_{ik} \geq 2\sum_{i=1}^{N-1} l''_{ik}$. Lemma **??** implies that if $j < k$, then $\sum_{i=1}^{N-1} l''_{ij} \geq \sum_{i=1}^{N-1} l''_{ik}$. In particular, this inequality holds when $j = k - .5$. Using inequality (**??**), we conclude that

$$\sum_{i=1}^{N-1} l'_{ik} \leq \sum_{i=1}^{N-1} l''_{ik} \leq 2\sum_{i=1}^{N-1} \sum_{k=1}^{N} l''_{ik},$$

which is the desired result. ■

Let us further specialize the model and assume $v_i - v_{i-1} \equiv \nu$, where ν is a positive constant and $c(i,j) = d(j-i)$ depends on the difference between the reported state and the true state.

Proposition 8 *Suppose $v_k - v_{k-1} > c(k-1,k)$ for all k and $v_i - v_{i-1} \equiv \nu > 0$ for all i. The probability of an honest report is a non-increasing function of the observed value.*

Proof. It suffices to show that if $j > i$ and type (j,t) dishonestly reports k, then type (i,t) prefers to report $k-j+i$ rather than to tell the truth. That is, if

$$W_k - d(k-j) - t \geq v_j \text{ implies } W_{k-j+i} - d(k-j+i-i) - t \geq v_i. \quad (24)$$

Implication (**??**) follows because

$$W_k - v_j = v_k - v_j + \alpha r_k = v_{k-j+i} - v_i + \alpha r_k \geq v_{k-j+i} - v_i + \alpha r_{k-j+i} = W_{k-j+i} - v_i,$$

where the first and last equations are definitions, the second inequality follows because $v_k - v_j = (k-j)\nu = v_{k-j+i} - v_i$ and the inequality follows from Lemma **??**. ■

We conclude this section with a discussion of related papers.

Abeler, Nosenzo, and Raymond **?** discuss a variety of models for an environment in which there are two possible outcomes and two possible reports. The models include preferences in which making a dishonest report lowers utility and in which agents' utility is increasing in their reputation for honesty, a force that operates similarly to our likelihood dimension. They find models that include both of these features help organize experimental data. The model of Abeler, Nosenzo, and Raymond **?** permits underreporting in equilibrium. Underreporting is ruled out in our model in equilibrium by Proposition **??**.[14]

Dufwenberg and Dufwenberg **?** study a model in which the agent's preference depends on the claim and a term similar to our social identity term. The term is proportional to an observer's expectation of the difference (if positive) between the claim and the observed value. In our model, the likelihood term is proportional to the probability that the report is honest. Hence, in their model, the social identity cost of making a dishonest report may depend on the level of dishonesty.[15] There is at least one important implication

[14] Utikal and Fischbacher **?** find evidence of under reporting. Utikal and Fischbacher find that no one reports the two (out of six) highest outcomes. Their sample size is small (twelve observations) and their subject pool (nuns) may be atypical.

[15] Dufwenberg and Dufwenberg **?** assume that agents have identical preferences and there is no direct cost of lying. It our notation, they assume $C(\cdot) \equiv 0$.

of this difference in assumptions: Dufwenberg and Dufwenberg show by example that Propositions **??** and **??** do not hold in their model. Their model typically exhibits multiple, qualitatively different equilibria, although they provide reasons to focus on a particular equilibrium. Similar to us, Dufwenberg and Dufwenberg show that partial lies are possible in equilibrium.

Khalmetski and Sliwka **?** analyze a special case of our model in which $c(i, j) = 0$ (they also assume that the prior distribution is uniform). This specialization permits them to fully characterize the unique symmetric equilibrium of the model. Whereas they focus on comparative statics with respect to the value of reputation (α), they identify some of the same important qualitative properties that we do.

In Dufwenberg and Dufwenberg **?** and Khalmetski and Sliwka **?** there is indifference across many possible lies. For this reason, randomization is essential in the construction of equilibrium. If k and k' are claimed dishonestly in equilibrium, then anyone who claims k dishonestly in equilibrium will obtain the same utility by claiming k'. In our most general specification, if $k \neq k'$, then type (k, t) may strictly prefer to make a different dishonest report than type (k', t). Nevertheless, our assumptions do imply that if two agents observe the same k and make different dishonest reports, then they must be indifferent between these reports. That is, type (k, t) and type (k, t') have the same preferences over lies for every k (although if $t' > t$, then (k, t) might prefer to report honestly when (k, t') prefers to lie). This observation follows from the assumption that the fixed cost of lying enters additively in $C(\cdot)$.

Garbarino, Slonim, and Villeval **?** analyze a model in which there are two possible observations and argue that reference dependence and loss aversion may influence lying behavior. In particular, they show that agents are more likely to dishonestly report the good outcome when the prior probability of the good outcome increases. This result is similar to our findings on the effect of varying the prior distribution of observed outcomes.[16]

[16] Unlike Abeler, Nosenzo, and Raymond **?**, Dufwenberg and Dufwenberg **?**, Khalmetski and Sliwka **?**, and this paper Garbarino, Slonim, and Villeval do not assume that preferences depend on beliefs about the honesty of a report. Garbarino, Slonim, and Villeval run double-blind experiments and argue that the social identity effect that we discuss should not influence payoffs when subjects know they are not being observed. We do not find this argument convincing for three reasons. First, our data allows us to compare observed to non-observed games. Observability matters in a way that loss aversion alone does not capture. Second, our model permits the social identity motive to be internal. Doing some-

4 Hypotheses

This section describes testable implications of the theory. The results section below will be based on these hypotheses.

Hypothesis 1 *If some type lies when reporting k, then no type with true value k lies.*

Hypothesis **??** is a consequence of Proposition **??**.

Hypothesis 2 *No agent underreports.*

Hypothesis **??** is a consequence of Proposition **??**.

Hypothesis 3 *The highest claim is made dishonestly with positive probability.*

Hypothesis **??** is a consequence of Proposition **??**.

Hypothesis 4 *When social identity concerns are strong enough, some agents will lie partially.*

Hypothesis **??** is a consequence of Proposition **??**.

Hypothesis 5 *There exists a threshold of true values, below which there are lies with positive probability, above which there are no lies.*

Hypothesis **??** is a consequence of Proposition **??**.

It follows from Hypotheses **??**–**??** and **??** that we should observe the number of reports of a k to be below the actual number of times the true value is k up to a cutoff for $k \leq k^*$ and the number of reports of k to be above the true number for $k > k^*$.

Hypothesis 6 *An increase in the probability that the true type is less than the highest type increases the number of values reported dishonestly and the probability of partial lies.*

---

thing that would appear dishonest even if she knows no one would find out may be costly to an agent. Her motive is to reinforce her social identity as an honest person. Third, agents may mistakenly believe others know what they know.

Hypothesis **??** is a consequence of Proposition **??**.

The next hypothesis refers to the exercise of increasing the number of states discussed in Proposition **??**. That is, we compare equilibria of a situation in which there are N equally likely, equally spaced, outcomes to one in which there are $2N$ equally likely, equally spaced, outcomes (with new outcomes inserted between old ones).

Hypothesis 7 *Increasing the number of states increases the range of the values that are reported dishonestly and increases the probability of partial lies.*

Hypothesis **??** states that splitting states lowers the cutoff that determines the lowest claim that would be made dishonestly. It is a consequence of Proposition **??**.

Hypothesis 8 *The lower the true value, the higher the fraction of dishonest reports.*

Hypothesis **??** is a consequence of Proposition **??**.

5 Experimental Design and Procedure

To test the theoretical predictions of our model regarding the dimensions of lying costs, we introduce two types of games, which we call **observed** and **non-observed** games.

5.1 Observed Games

The observed game is a variation of a cheating game, in which we can observe the individual lying behavior. In this game, we ask participants to click, in private, on one of ten boxes on a computer and reveal an outcome. We use three different observed game variations. In the "Numbers" treatment, the outcomes behind the ten boxes are numbers between one and ten, where each box has a different number.[17] After seeing the number, the participant is asked to report it to the experimenter, knowing that payments will be

[17]Both the subject and the experimenter know that the numbers vary between one and ten. Neither our model nor the experiments consider what happens in asymmetric-information cases in which the subject has private information regarding the outcome

equal to the number s/he reports in Euros. In this treatment, we know how often and to what extent participants lie because we can later observe the actual number each participant saw and compare it with the number s/he reported.

As discussed earlier, lying costs may depend on the distance between what the subject observes and what the subject says (outcome dimension). In the Numbers treatment, there is a natural and common measure of the distance between observed outcomes and reports. For example, if the participant observes a “four” and reports “ten,” her report is six units from the truth. One can imagine that lying costs depend on this distance.

Another possibility is that lying costs depend on the payoff gained by the report relative to what the participant would earn if she reports honestly (payoff dimension). In the Numbers treatment, payoffs are linked directly to outcomes so one cannot distinguish the outcome dimension from the payoff dimension. The second treatment, “Numbers Mixed,” is designed to separate the reported outcome dimension from the payoff dimension. This treatment is similar to the Numbers treatment, but the ten numbers are assigned to the ten payoffs in a random order. Table A1 presents the assignment we defined in a random draw.

No natural way to measure the distance between different observed outcomes exists in the Numbers Mixed treatment. In the third observed treatment, “Words,” there is no natural ordering of the outcomes independent of payoffs. In this treatment, participants are asked to click on one of ten boxes in private and are told that the outcomes behind the boxes are ten Lithuanian words; each box has a different word. The words have payoffs between one and ten Euros assigned to them, as presented in Table A2 (in Appendix A).

There is no natural notion of distance in the outcome dimension because none of the participants knows Lithuanian and the words appear to be similar six-letter strings. Participants may distinguish between reporting truthfully or not, but we assume that the “outcome cost” of reporting (the Lithuanian word) “stirna” when the outcome is “vilkas” is the same as the outcome cost of reporting “kiskis” when the outcome is “vilkas.” More generally, we

space itself. Consider, for example, a game in which it is common knowledge that the observer believes the die is a standard six-sided die, but the subject knows in fact that the die has no six. Would the agent be more or less likely to lie in such a treatment than in the baseline setting with symmetric information? We leave testing of this situation for future research.

assume that the outcome cost is zero for honest reports and the same for all dishonest reports.

5.2 Non-Observed Games

As discussed in Sections **??** and **??**, a participant might care about how she is perceived. When the experimenter does not observe the outcomes, the participant may refrain from reporting the highest number to signal she is not a liar.

In the "Basic" non-observed treatment, we give the participant a sealed envelope with ten folded pieces of paper that have numbers from one to ten on them. We ask the participant to take out one piece of paper, observe the number she took out, put it back into the envelope, and then report it. As in the observed treatment, payments are equal to the number reported in Euros. However, in contrast to the observed game, the experimenter can never know the observed outcome. If social identity concerns affect the lying decision, we would expect more participants to lie and/or a higher fraction of participants to partially lie in the non-observed than in the observed game. Lying costs that depend only on payoffs and outcomes could not explain a difference in behavior between the observed and non-observed games.

The model predicts that decreasing the prior probability of the highest outcome affects the number of values reported dishonestly. In the "Low Probability" non-observed treatment, we use a similar procedure as in the Basic non-observed treatment and adjust the prior probabilities of the outcomes occurring. We give the participant a sealed envelope with 100 folded pieces of paper that have numbers from one to ten on them. We inform the participant that eleven pieces of paper have the number "one" on them, eleven pieces have the number "two" or them, and so on until "nine." However, one piece of paper has the number "ten" on it. As in the Basic non-observed treatment, the payments are equal to the number reported in Euros. Guided by the model's predictions, we expect subjects to report a higher range of values , because the chance of drawing a ten is only 1%, as opposed to 10% in the Basic non-observed treatment. If the prior distribution of outcomes influences reports, then outcome/payoff lying costs are insufficient to understand data.

In the final treatment –"100-States" non-observed treatment – we investigate the robustness of the predictions in the Low-Probability treatment. In this treatment, we give participants a sealed envelope with 100 pieces of paper with numbers between one and 100 on them and inform them that each piece of paper has one of the numbers on it. Participants receive the equivalent in Euros to the number they report divided by ten. Whereas the probability of drawing 100 is the same as in Low-Probability treatment, all the other outcomes are equally likely. This treatment allows us to test whether the partial lying that we might observe in the Low-Probability treatment is the result of the difference in the relative probability of the highest state and lower states or is due to the fact that the absolute probability of the highest state is low.

5.3 Experimental Procedure

We conducted the experiments between April 2015 and April 2016 at the Cologne Laboratory for Economic Research, University of Cologne. We used the experimental software zTree (Fischbacher **?**) and recruited participants via ORSEE (Greiner **?**). Overall, we recruited 916 participants (55.9% female), and none of them participated in more than one session. We collected 102–390 observations per treatment. Participants played only our treatment in the experimental session with a session lasting approximately 30 minutes. See Appendix C for instructions.

After being randomized into a treatment, participants read the instructions on the computer screen, and were allowed to ask questions privately. Then, depending on the treatment, participants either received the envelopes with numbers and were asked to pick one, or were asked to click on one of the boxes on the computer screen and reveal the number. After observing the number, participants reported the outcomes on a sheet of paper and filled out a post-experiment questionnaire that included questions on gender, age, field of study, and motives behind the decisions. At the end, participants privately received their payoffs in cash and left the laboratory. Table **??** presents all of our treatments and the number of participants in each.

6 Results

In what follows, we first present the data from the observed treatments and then move to testing the hypotheses that are based only on the observed treatments (Hypotheses **??**–**??**, **??**, and **??**). We then report the results from the non-observed treatments and the tests of the corresponding hypotheses

FiguresTables_jpg/Table-1.jpg

(Hypotheses **??** and **??**–**??**).

6.1 Observed Game

We divide our discussion into two subsections. We first describe the descriptive statistics and then discuss the relationship between behavior and the hypotheses.

FiguresTables_jpg/Figure-1.jpg

6.1.1 Outcome and Payoff Dimensions in the Observed Game

Figure **??** presents the distributions of actual and reported payoffs in the observed game and is the first indicator that the reported payoffs are higher than the actual payoffs resulting from the outcomes (numbers or words).

We find that in all the observed treatments the reported numbers are significantly higher than the observed outcomes with $p < 0.001$ (Wilcoxon

Matched-Pairs Signed-Ranks Test).[18] Overall, 26%, 33%, and 27%, participants lie in the Numbers, Words, and Numbers Mixed treatments, respectively. The overall level of lying is not significantly different between the treatments in pairwise comparisons in a Fisher exact test ($p > 0.1$).

The observed game also allows us to analyze the probability of lying conditional on the actual payoff observed, as presented in Figure **??**. We can see from the figure that in the Numbers treatment, the lower the observed outcome is, the more likely participants are to lie. For example, only 5% of participants who observed a nine overreport their number, whereas 47% of the participants who observed a one did. The results show a significant negative correlation (Spearman's rho = -0.318, $p < 0.001$) between the payoff observed and the probability of lying. In the Words treatment, we also observe a negative correlation, but less strong than in the Numbers treatment (Spearman's rho = -0.202, $p = 0.042$). In the Numbers Mixed treatment, a marginally significant correlation exists between the actual payoff and the probability of lying (Spearman's rho = -0.170, $p = 0.079$).[19]

Another important feature of the observed treatment is that it allows us to know what payoff people who lie report. Figure **??** presents these data. We find no correlation between the actual and reported payoff when one lies (Spearman's rho = 0.052, $p = 0.601$ for the Numbers treatment; rho = 0.188, $p = 0.288$ for the Words treatment and rho= -0.021, $p = 0.914$ for the Numbers Mixed treatment). This finding is consistent with models of Dufwenberg and Dufwenberg **?** and Khalmetski and Sliwka **?** that assume no variable costs of lying.[20]

Although the actual observed payoffs are not different between the treatments (i.e., the randomization worked; $p > 0.1$, MWU), the average reported payoff in the Numbers treatment is only marginally lower than in the Words treatment (7.02 versus 7.39, respectively; $p = 0.090$, MWU), and is not statistically different from the average reported payoff of 7.03 in the Numbers

[18] All tests in the paper are two sided. We call an effect highly significant, significant, or marginally significant if the test generates $p < 0.01$, $p < 0.05$, $p < 0.1$, respectively.

[19] One person under-reported in this treatment (observing ten and reporting four) and two did not click on any boxes and then reported a ten. The two participants who did not click are excluded in Figure **??** and in the Spearman's correlations, because they have no observed outcome. If the two participants are not excluded, Spearman's rho amounts to -0.213 with $p = 0.025$.

[20] If $c(i, k) < c(i, j) + c(j, k)$ in our model, then, conditional on lying, lower types would tell bigger lies than higher types.

FiguresTables_jpg/Figure-2.jpg

Mixed treatment ($p = 0.198$, MWU). The difference between the Numbers Mixed and Words treatments is also not statistically significant ($p = 0.205$, MWU). Thus, in the extensive margin, the reporting is not significantly different between the observed treatments.[21]

Next, we consider the payoffs reported by participants who lie, which are

[21] The extensive margin corresponds to the fraction of people who lie; the intensive margin corresponds to the size of the lie for people who choose to do so.

FiguresTables_jpg/Figure-3.jpg

presented in Figures ?? and A1. We observe that in the Numbers treatment, 68% of participants who lie, lie to the full extent by saying ten. This fraction is 91% in the Words treatment and 80% in the Numbers Mixed treatment. The fraction of participants who lie by reporting ten in the Words treatment is significantly higher than in the Numbers treatment ($p = 0.007$, Fisher exact test). The average payoff reported by participants who lie in the Words treatment, 9.80, is also significantly higher than the Numbers treatment (9.80 versus 9.32, respectively; $p = 0.011$, MWU). Thus, in the intensive margin,

FiguresTables_jpg/Figure-4.jpg

we find significant differences between lying behavior in the Words and Numbers treatments.

The difference between lying in the Words and the Numbers treatments suggests that the outcome dimension affects lying costs, because we observe less partial lying in the Words treatment without a clear notion of partial lying on the outcome dimension than in the Numbers treatment in which there is an intuitive notion of intermediate lies. This finding suggests that some participants perceive reporting “eight” when observing “four” a smaller lie than reporting "ten" in the Number treatment, whereas in the Words treatment, reporting "alyvos" when observing "vilkas" has the same outcome cost as reporting "alyvos" when observing "stirna." However, the role of the outcome dimension on the extent of lying is relatively small, because only 8.45% of the participants (33 out of 390) lie partially in the Numbers treatment and the fraction decreases to 2.94% (3 out of 102) in the Words treatment. In addition, as we show above, no significant effect on the extensive margin exists.

The absence of partial lying in the Words treatment suggests that the payoff dimension has no effect on the cost of lying on the intensive margin. When observing an outcome that results in 4 Euros if reported honestly, the cost of dishonestly reporting something that leads to a payoff of 6 Euros does not appear to be significantly lower than the cost of dishonestly reporting something that leads to a payoff of 8 Euros.

In the Numbers Mixed treatment, the fraction of participants who lie is between the Numbers and Words treatments and is not significantly different from the two treatments ($p = 0.257$ and $p = 0.285$, Fisher exact test). The average payoff reported by participants who lie in the Number Mixed treatment is not significantly different from either the Words or the Numbers treatments (9.80 versus 9.32, respectively; $p = 0.205$ and $p = 0.250$, MWU).

Figure **??** shows the results for the Numbers Mixed treatment with respect to the outcome dimension instead of the payoff dimension. The results clearly show that the lying behavior in this treatment is not related to the outcome dimension – the decision to lie does not depend on the actual number observed (see Figure 5b; Spearman's rho = -0.157, $p = 0.102$) and when participants lie, they lie mostly by reporting a "two," which results in a payoff of ten (Figures 5a and 5c). That is, when there is a trade-off between the outcome and payoff dimensions, the participants lie according to the payoffs and neglect the outcome dimension. This finding again suggests only a limited effect of the outcome dimension on the lying cost.[22]

Based on the comparisons between the observed treatments, we conclude that the outcome dimension has a limited effect on the intensive margin and no effect on the extensive margin. We also conclude that the payoff dimension

[22]We do not have a test of what happens if we change the payoffs associated with the decisions. For this reason, we cannot estimate the effect of the payoff costs on the extensive margin. Providing such a test is not trivial, because changing the payoffs would lead to changes in the behavior independent of the intrinsic lying cost (see Kajackaite and Gneezy **?**). We leave this exercise for future research.

FiguresTables_jpg/Figure-5.jpg

has no effect on the lying cost on the intensive margin.

6.1.2 Hypothesis Testing

The results from the observed treatments allow us to test Hypothesis **??**, which states that if a participant observing k ever lies, then no one ever lies by saying k. As Figure **??** and Table A3 show, the results are generally consistent with this prediction. We use the pooled data from the observed game to count for deviations from the prediction with respect to the theory; we define a behavior as a "mistake" if it violates Hypothesis **??** (i.e., if given that someone observing k lies, someone lies by reporting a k). Under this definition, mistakes are the minimum of the fraction who lie when observing the number and the fraction who lie by reporting the number. In Figure **??**, the mistakes are marked with a dashed line. Lines 1–3 in Table A3 (left side of the Figure) contain no mistakes. Line 4 in Table A3 contains one mistake: participants lie after observing a four and one participant out of 167 (0.60%) lies by reporting four. Lines 5, 6, 7, 8, and 9 contain 2, 5, 6, 13, and 15 mistakes, respectively (1.20%, 2.99%, 3.59%, 7.78%, and 8.98% out of 167 participants). Line 10 contains one mistake, because one person out of 73 (1.37 %) observing a ten lies downward.

Overall, we observe 43 mistakes for 602 participants (7.14%) in our data. We conclude:[23]

Result 1 *The data show that if a participant observing k lies, then only a small fraction of participants lie by saying k.*

Hypothesis **??** states that a participant would not underreport her payoff. This hypothesis is easy to test in our data. As we reported above, only one participant out of 602 underreported in our experiment. Therefore, we conclude:

Result 2 *Most participants (99.83%) do not underreport their payoffs.*

Hypothesis **??**, which asserts that the highest claim is made dishonestly with a positive probability, is also clearly supported by the data presented in Figure **??**. In particular, we observe that of the people who lie, 68%, 80% and 91% of participants report the highest possible payoff in the observed treatments:

Result 3 *Of the participants who lie, a high fraction (an average of 74.85%) report the highest payoff.*

We find partial support for Hypothesis **??**, that there exists a threshold of actual payoffs, below which there are lies with positive probability and above which there are no lies. In particular, in the observed treatments 27.63-46.94% (an average of 34.26%) of participants lie when observing a payoff below nine, but only 9.68% and 1.37% lie after observing nine or ten, respectively (see Figure **??**, Figure **??**, and Table A3).

[23]We number the Results in parallel to the Propositions and Hypotheses, but we report them in a slightly different order. We hope that the reader will not be too alarmed to find that we report Result 5 immediately after Result 3. The patient reader will find Result 4.

FiguresTables_jpg/Figure-6.jpg

Result 5 *There is a threshold of actual payoffs of nine, below which there is a high fraction of lies (average 34.26%), and above which there are only few lies (average of 5.19%).*

We report support for Hypothesis **??**, which states that the lower the true value, the higher the fraction of dishonest reports, in Figure **??**. As we have observed, there is a significant negative correlation between the payoff observed and the probability of lying in the observed game:

Result 8 *There is a significant negative correlation between the payoff observed and the probability of lying in the observed game: The lower the actual payoff, the higher the fraction of dishonest reports.*

6.2 Non-observed Game

Figure **??** presents the results from the non-observed treatments. The graphs show the distributions of the reported payoffs on the aggregate level and, as in the case of the observed treatments, indicate (this time only statistically) that reported payoffs are higher than expected observed payoffs; a Kolmogorov-Smirnov test confirms participants lie significantly in the non-observed treatments ($p < 0.001$).

Whereas 14% report a nine in the Numbers observed treatment (which is not significantly higher than the actual fraction of 11% who observed a nine; $p = 0.106$, binomial test), 22% report a nine in the Basic non-observed treatment, which is significantly more than the theoretical prediction ($p < 0.001$). The difference between the Basic and Numbers treatments in reporting a nine is significant ($p = 0.033$, Fisher test). That is, in the non-observed treatment, some of the participants who lie do not report the maximal payoff. This result is predicted by Proposition **??** and Hypothesis **??**, assuming that some participants care about their social identity. In the non-observed treatment, to signal to the experimenter that she does not lie, a participant who lies may choose to claim high but not maximal numbers, such as eight or nine. Also supporting the social-identity-concern prediction, we find that the overall level of lying is higher in the Basic non-observed treatment than in the Numbers observed treatment (average reported number/payoff of 7.81 vs. 7.02, respectively, $p = 0.016$, MWU), indicating some participants do not lie in the observed treatment because the experimenter may know they lied.[24] Hence, we conclude:

[24] In the observed game, a lie must lead to $\rho = 0$. Therefore, the incentive to tell a partial lie in the observed game is lower (in the non-observed game, someone might tell a partial lie in order to improve her social identity relative to a full lie). Even without variation in ρ, partial lies are possible in the observed game (due to variations in the cost function), but observability makes all lies less important. Social identity does play a role in the observed game because the social identity term adds α to honest reports (and 0 otherwise).

FiguresTables_jpg/Figure-7.jpg

Result 4 *In line with social identity concerns, a larger fraction of participants lies partially in the non-observed treatment than in the observed treatment.*

Previous evidence showing partial lying in variations of the non-observed treatment was interpreted as a desire to maintain a positive self-image. Most notably, Mazar, On, and Ariely **?**, conclude that "A little bit of dishonesty gives a taste of profit without spoiling a positive self-view." Although replicating the partial-lying finding of Mazar et al. and others, our results do not support their interpretation. Instead, our results suggest that partial lies are primarily due to social identity concerns, because the partial lying that might be caused by self-image concerns is low in the observed games and substantially increases in the non-observed game.

Hypothesis **??** predicts that a lower prior probability of the highest outcome will increase the number of values reported dishonestly. The Low-Probability and 100–States non-observed treatments were designed to test this prediction. Recall that in the Low-Probability treatment, we reduced the probability of a "ten" to 1%. Figure 7b presents the results from the Low-Probability treatment.

Consistent with the model's predictions, lowering the prior of the highest outcome increased the range of values reported dishonestly. Although in the Basic treatment, only the fractions of reports of nine and ten are higher than the prior (22% and 37% compared to 10% prior; $p < 0.001$, binomial test), in the Low-Probability treatment, participants overreport eight, nine and ten (marginally) statistically significantly and seven is overreported but not significantly so. Here, 16% report a seven (compared to 11% prior, $p = 0.120$), 17% an eight (compared to 11% prior, $p = 0.063$), 34% a nine (compared to 11% prior, $p < 0.001$) and 6% a ten (compared to 1% prior, $p < 0.001$). That is:

Result 6a *Consistent with Hypothesis* **??***, when a payoff of 10 has a 1% chance, a larger number of values is reported dishonestly.*

Hypothesis **??** predicts that a lower prior probability of the highest outcome will increase the probability of partial lies. In the Basic non-observed treatment, only nine and ten are overreported relative to the expected fraction. Twenty three (22.33%) out of 103 participants reported nine compared with the expected 10% who observed nine. Thus, we estimate the partial lying to be 12.33% in this treatment. In the Low-Probability non-observed treatment, seven, eight, nine, and ten are overreported. 11% are expected to observe each of the numbers seven, eight and nine, but 17 (15.89%), 18 (16.82%) and 36 (33.64%) out of 107 participants claimed to do so. Therefore, in the Low-Probability treatment, the estimated partial lying is 33.36% $(71 - 35.31)/107$, with 71 being the sum of participants claiming 7–9, and 35.3, the expected fraction of participants who observed 7–9.

Result 6b *Consistent with Hypothesis* **??***, when a payoff of ten has a 1% chance, the fraction of partial lies increases relative to when the payoff of ten has a 10% chance.*

López-Pérez and Spiegelman **?** conduct an experiment related to this result. In their experiment, a subject sees either the color green or blue and must report a color. She receives a higher monetary payoff for reporting green (the monetary payoff depends only on the report, not on the observed color). The authors also ask subjects to state the fraction of subjects that report green dishonestly and also ask subjects to estimate the average assessment of dishonest reports provided by previous subjects. The authors run two treatments, which vary the prior probability that the true state is green. López-Pérez and Spiegelman point out that if the cost of lying depends only on the report, then the amount of lying should be the same in both treatments. They show that most subjects either report honestly or always report green. Increasing the prior probability that the true color is blue decreases the probability that subjects report blue independent of the true color, although this finding is not statistically significant. This result is consistent with our Result **??**.[25]

The low-probability treatment demonstrates that making the highest possible outcome relatively less likely than the other outcomes makes participants report a larger range of outcomes dishonestly. In the 100-States treatment we test whether lowering the absolute probability of the highest state has a similar effect. Figure 7c presents the results and Figure 7d presents the aggregate results.

To test Hypothesis **??** – that increasing the states increases the number of values reported dishonestly – we compare the data from the 100-States treatment with the Basic treatment. As described above, in the Basic treatment, only nine and ten are overreported (two out of ten possible outcomes). In the 100-States treatment, we find that 22 out of 100 outcomes are overreported relative to the expected 1%. The cutoff at which numbers are reported dishonestly is lower than in the 10-States condition. Significant overreporting starts at 62 (out of 100) in the 100-States condition. It starts at nine (out of ten) in the 10-States condition. We summarize this finding in Result **??**.

Result 7a *Consistent with Hypothesis* **??**, *the range of values reported dishonestly in the 100-States treatment is larger than in the 10-States treatment.*

[25] López-Pérez and Spiegelman present a theory that predicts that increasing the prior probability of the low-value observation (blue) will increase the fraction of subjects who report green when they observe blue. This prediction is the opposite of what we find. In their model, agents suffer losses from lying to the extent that lying is perceived as unusual. When the prior probability of green is low, the conditional probability that a report of green is dishonest is high, which lowers the cost of the lie in their model.

Finally, to test Hypothesis ?? – that increasing the number of states will increase the probability of partial lies – we compare estimated partial lying in the Basic treatment (12.33%) with the 100-States treatment. We estimate partial lying in 100-States treatment by analyzing pooled intervals and identifying which intervals are overreported. We divide the data into 11 groups: 1–9, 10–19, ..., 90–99, and 100. We find that 60–69, 80–89, and 90–99 are overreported, with 60.58% reporting those outcomes. We conclude that partial lying amounts to 30.58% (60.58−30) in the 100-States treatment.

Result 7b *Consistent with Hypothesis* **??**, *increasing the states increases the fraction of partial lies.*

To obtain Result **??**, we isolate the highest state and pooled lower intervals in groups of ten. Isolating the highest state is consistent with our objective of identifying whether reducing the probability of the highest state leads to a lower fraction of agents making the highest claim. If we pool together states 1–10, 11–20, ..., 91–100, then there is statistically significant overreporting only in the two highest pools, which is consistent with Proposition **??** and, in particular, the ideas that increasing the number of states does not qualitatively change lying behavior.

7 Conclusion

In this paper, we formalize an important aspect of the intrinsic costs of lying – how the size of the lie affects the decision to lie. We discuss three possible kinds of lying costs: a cost related to the distance between the true outcome and what is reported; a cost related to the monetary gains generated by the lie; and a cost associated with the probability that a statement is perceived to be dishonest. Although the literature has discussed the first two dimensions, it has neglected the third dimension.

The model we construct allows us to consider the influence of the size of the lie on lying decisions and to generate novel predictions that we test experimentally. In line with the properties of the equilibrium of our model, we find evidence for a cutoff value: If the payoff associated with the observed outcome is high enough, then people do not lie, and lies occur only when the payoff is below this cutoff. In equilibrium, subjects who make higher claims are perceived as being less honest. In the experiment, as the model predicts, dishonest claims of the maximal value always occur. When lying does not lead to the lowest possible social identity (that is, by making the outcome non-observed), there is more lying and, in particular, more partial lying. Another finding that supports the social identity argument is that when making the maximal outcome less likely ex ante, the frequency of partial lies increases.

We conclude that social identity has an important impact on lying costs. Our findings indicate that the other two dimensions – the outcome and the payoff dimensions– have smaller effects on lying behavior. For example, we find that the outcome dimension has no effect on the number of people who choose to lie and a small effect on partial lying.

Our paper offers a formal treatment and systematic experimental analyses of the intrinsic cost of lying and its interaction with the size of the lie. Several interesting experiments follow from our discussion. We hope they will be the subject of future research. A natural extension is to test what happens when we introduce a Numbers Mixed and a Words treatments, designed to separate the reported outcome dimension from the payoff dimension, into the non-observed game. A more challenging problem is to estimate the effect of the payoff costs on the extensive margin. Another interesting question is what happens when there is an exchange rate between the benefit to the decision maker and the cost to another player. Although we have tried to understand the importance of the interplay between social identity and regular interpretation of reputation, open questions remain. For example, participants may believe their reports influence the chance they will be invited back to the laboratory, causing them to view the interaction as more than a one-shot game. Future research can use a double-blind procedure in which the experimenter does not know the decision made by individual participants.