

# Disguising Lies - Image Concerns and Partial Lying in Cheating Games\*

Kiryl Khalmetski and Dirk Sliwka<sup>†</sup>

September 17, 2019

## Abstract

We study equilibrium reporting behavior in cheating games when agents have a fixed cost of lying and image concerns not to be perceived as a liar. We show that equilibria naturally arise in which agents with low costs of lying randomize among a set of the highest potential reports. Such equilibria induce a distribution of reports in line with observed experimental patterns. We also find that higher image concerns lead to an increase in the range of reported lies while the effect of the fixed cost of lying is the opposite.

**Keywords:** cost of lying, image concerns, cheating game, truth-telling, deception.

**JEL codes:** D82, D83, C72.

---

\*Accepted for publication in *American Economic Journal: Microeconomics*.

<sup>†</sup>Khalmetski: Faculty of Management, Economics and Social Sciences, University of Cologne, Albertus-Magnus-Platz, 50923 Cologne, Germany (e-mail: kiryl.khalmetski@uni-koeln.de). Sliwka: Faculty of Management, Economics and Social Sciences, University of Cologne, Albertus-Magnus-Platz, 50923 Cologne, Germany (e-mail: sliwka@wiso.uni-koeln.de). We thank three anonymous referees, Johannes Abeler, Martin Dufwenberg and Joel Sobel for very helpful comments and suggestions. Financial support of the German Research Foundation (DFG) through the Research Unit “Design and Behavior” (FOR 1371) is gratefully acknowledged.

# 1 Introduction

Economists have only recently started to experimentally investigate lying behavior. Gneezy (2005) has shown that a significant share of subjects have a strong aversion to lying besides purely distributional preferences.<sup>1</sup> Subsequent studies explored various determinants of individual lying aversion such as the payoff consequences for the sender and the receiver of the message (Erat and Gneezy, 2012), the form of communication (Lundquist et al., 2009), or anticipated trust of the receiver (Charness and Dufwenberg, 2006; Sutter, 2009).

An important experimental paradigm for studying motivations for honesty has been developed by Fischbacher and Föllmi-Heusi (2013) which already in the rather short time frame since its publication has been used in a large number of experimental studies. A recent meta-study by Abeler, Nosenzo and Raymond (forthcoming) already includes more than 70 studies using this paradigm. In this setting, subjects privately have to roll a die and then report the number they have observed. The payoff is simply equivalent to the reported number (except for the number 6 which leads to a payoff of zero). An important advantage of the method is that it avoids strategic interactions between subjects and thus facilitates the identification of intrinsic motives for lying behavior. Moreover, in the baseline design the experimenter cannot verify the reported messages on the individual level, hence, subjects can be close to certain that lies are not discovered, but lying behavior can still be inferred from the distribution of reported numbers. If all subjects told the truth each number should be reported with a likelihood of  $1/6$ . However, in the experiment the fractions of numbers 4 and 5 among all reported numbers were significantly above the expected  $1/6$  (i.e., 27.2 percent for the 4 and 35.0 percent for the 5). Hence, a substantial number of subjects did not tell the truth - but also apparently many did not maximize their own material payoff (i.e., did not report the number 5). As a possible explanation, Fischbacher and Föllmi-Heusi (2013) suggest that some subjects might have a reputational concern motivating them to disguise their lying to an outside observer (which can be an experimenter or a future self) by reporting not the payoff-maximizing number. This conjecture was further substantiated by the experimental evidence that partial lying in cheating games turns to be significantly less frequent if lying can be individually verified by the experimenter (Gneezy, Kajackaite and Sobel, 2018; Abeler, Nosenzo and Raymond, forthcoming).

We now formalize this idea by developing a model which assumes both a

---

<sup>1</sup>Correspondingly, people have also strong aversion to being lied to as shown by Brandts and Charness (2003).

fixed cost of lying *and* social image concerns not to be viewed as a liar and study whether equilibrium behavior can indeed predict patterns such as those described in the experiment. In our model agents consider three motives: (i) they are better off when earning more money, (ii) they have an intrinsic cost of lying (which varies between individuals) and (iii) an agent’s utility is decreasing in the likelihood with which an outside observer believes her to have lied after having observed the reported number.<sup>2</sup> We show that such context naturally gives rise to equilibria in which agents who lie randomize their reporting behavior among a subset of the most favorable reports. The results remain robust to adding an additional component to the intrinsic cost of lying which linearly depends on the size of the lie.

We furthermore show that while stronger social image concerns reduce the likelihood of lying, they expand the range of reports that agents lie to. If image concerns are weak, all liars report the highest feasible number. In the opposite case when image concerns become very strong, liars randomize over the whole set of possible reports excluding the smallest number. Thus, too strong reputational concerns of the sender may actually backfire for the receiver since then she can never be sure to have obtained the correct information (except for the case of the lowest message). Notably, we show that this effect goes in the opposite direction of the effect of an increase in the fixed costs of lying, which leads to a smaller range of reported lies. Finally, we consider the effect of the stake size. While the total *rate* of lying unambiguously increases with the stake size, the *range* of reported lies can both increase and decrease depending on the parameters. The latter fact is due to the opposing effects from a relative decrease in the reputational concerns and the fixed cost of lying.

With very few calibrated parameters, the model produces a good fit of the predicted behavior to the experimental data from Fischbacher and Föllmi-Heusi (2013) and related experiments. While our model is primarily motivated to explain the empirical evidence from experimental cheating games, it is worth noting that such games also capture structural properties of many real-life communication settings where: 1) transmitted information is not easily verifiable (or verifi-

---

<sup>2</sup>Models of belief-independent costs of lying were developed by Kartik, Ottaviani and Squintani (2007) and Kartik (2009). Ottaviani and Sørensen (2006*a*) and Ottaviani and Sørensen (2006*b*) considered settings where the sender cares about the inferred precision of his private signal. The concern for the inferred social preferences was modeled in Bernheim (1994), Bénabou and Tirole (2006) and Ellingsen and Johannesson (2008) where an agent prefers to be perceived as having a certain *type* of preferences. In contrast to these models, we assume that the sender cares about the receiver’s inference regarding the *action* of lying. See Abeler, Nosenzo and Raymond (forthcoming) for further classification of models relevant for the considered setting, including the models of inequality aversion, social conformity and guilt aversion.

able with a significant delay), 2) there are material incentives to misreport, and 3) the sender of the message might care about not being perceived as a liar at the time of communication. Potential examples include medical diagnoses, promises in political election campaigns, the sale of credence goods, and academic reference letters.

Independently from our work, a few recent papers aim to explain the experimental data from cheating games. In Gneezy, Kajackaite and Sobel (2018) and Abeler, Nosenzo and Raymond (forthcoming), the main model specifications also combine both intrinsic costs of lying and image concerns depending on the probability others assign that the agent lies.<sup>3</sup> Unlike us, Dufwenberg and Dufwenberg (2018) incorporate only (belief-dependent) costs that are proportional to the size of a lie perceived by the receiver. In terms of qualitative analytical results, the closest study to ours is of Gneezy, Kajackaite and Sobel (2018) who showed some of the important equilibrium properties also outlined in our analysis. While our model has somewhat stricter restrictions on preferences (such as permitting only linear dependency of the intrinsic lying cost on the size of the lie), at the same time this allows us to provide a complete mathematical characterization of the (unique) equilibrium distribution of reports. Moreover, our modeling framework allows us to precisely characterize the comparative statics of the equilibrium with respect to an increase in the image concerns, fixed lying costs and monetary stakes, obtaining qualitatively different and testable predictions for each case that can explain several previous experimental findings.<sup>4</sup>

The remainder of the paper is organized as follows. Section 2 presents the model setting, Section 3 provides equilibrium analysis, Section 4 derives comparative statics, Section 5 shows the results of the calibration of the model to the experimental data from Fischbacher and Föllmi-Heusi (2013) and Gneezy, Kajackaite and Sobel (2018), and Section 6 concludes.

## 2 The Model

We consider a population of agents who play a variant of the Fischbacher and Föllmi-Heusi (2013) cheating game. First, an agent privately observes a uniformly distributed random (integer) number  $y \in \mathcal{Y} = \{0, 1, \dots, K\}$  with  $K > 1$  such that

---

<sup>3</sup>Garbarino, Slonim and Villeval (2016) use a reduced form approach to model reputational concerns in a cheating game (combined with intrinsic lying costs), while focusing on the effect of loss aversion in this setting.

<sup>4</sup>The focus of the comparative statics analysis in the other studies is somewhat different. In particular, Gneezy, Kajackaite and Sobel (2018) primarily consider the effect of the underlying distribution of the observed numbers.

$\Pr(y) = 1/(K + 1)$ .<sup>5</sup> She then has to report this number to a receiver making a report  $x \in \mathcal{Y}$  which leads to a payment equal to  $x$  to the agent. We assume that agents incur a personal fixed cost of lying  $l > 0$  such that their utility is reduced by  $l$  whenever they report  $x \neq y$ .

Besides, agents have a concern to be viewed as being honest. In particular, we assume that utility is decreasing in the likelihood that the receiver (or some outside observer) thinks that she has told a lie, i.e., in  $\Pr[y \neq x|x]$ .<sup>6</sup> Thus, her utility function is

$$u(l, x, y) = x - l \cdot I_{\mathcal{L}}(x, y) - \eta \cdot \Pr[y \neq x|x] \quad (1)$$

where  $\mathcal{L}$  denotes the set of lies, i.e.,  $\mathcal{L} = \{(x, y) \in \mathcal{Y}^2 : x \neq y\}$ ,  $I_{\mathcal{L}}$  is the corresponding indicator function, and  $\eta > 0$  captures the extent to which agents dislike being viewed as liars by the receiver. Since the utility function depends on the receiver's beliefs, we obtain a psychological game (Geanakoplos, Pearce and Stacchetti, 1989; Battigalli and Dufwenberg, 2009). The term  $-\eta \cdot \Pr[y \neq x|x]$  is referred below as the agent's reputational payoff.<sup>7</sup>

We assume that  $l$  is distributed according to a strictly increasing continuous cdf  $F(l)$  with  $F(l) = 0$  for  $l \leq 0$  and  $F(l) > 0$  for  $l > 0$ .<sup>8</sup> Note that our model allows for a large probability mass on arbitrarily small lying costs: The assumption that  $F(0) = 0$  just implies that agents who are otherwise indifferent between lying and telling the truth, will tell the truth (similarly to a lexicographic preference for truth-telling). We also assume that  $F(K + \eta) < 1$  which ensures that in any equilibrium there is a strictly positive (potentially very small) probability mass of agents who always report truthfully and, in turn, all numbers are reported with positive probability.<sup>9</sup>

<sup>5</sup>Setting the lowest number to 0 is without loss of generality for the qualitative results. The case of  $K = 1$  excludes overreporting of more than one number by construction, and hence is not subject of our analysis.

<sup>6</sup>See also Bénabou and Tirole (2006), Fischbacher and Föllmi-Heusi (2013) and Dufwenberg and Dufwenberg (2018) for discussions regarding the agent herself being a plausible audience.

<sup>7</sup>When  $\eta = 0$  an agent who decides to lie always lies to the full extent. Thus, a fixed lying cost (that does not depend on the size of the lie) cannot by itself explain overreporting of numbers strictly lower than  $K$  – the main empirical puzzle in the experiment of Fischbacher and Föllmi-Heusi (2013). While non-linear lying costs may explain some patterns, Abeler, Nosenzo and Raymond (forthcoming) show that any conventional model with just intrinsic lying costs would still be inconsistent with their experimental findings that 1) subjects lie substantially more towards the high state when the ex-ante probability of this state is higher, and 2) the distribution of reports significantly depends on their observability. They conclude that reputational concerns are the most plausible explanation for the observed patterns.

<sup>8</sup>Gneezy, Rockenbach and Serra-Garcia (2013) provide empirical evidence for heterogeneous fixed lying costs with a large share of subjects being characterized by intermediate costs.

<sup>9</sup>Indeed, the meta analysis by Abeler, Nosenzo and Raymond (forthcoming) shows that this

While we assume in the main body of the paper that agents have a fixed cost of lying, we show in Section 3.3 that our results are robust to incorporating a linear size-dependent cost of lying into the agent’s utility function.

### 3 Equilibrium Analysis

This section first derives general properties of the equilibrium reporting strategies, then provides a complete characterization of the unique equilibrium distribution of reports, and finally shows the robustness of the results to adding a linear size-dependent cost of lying.

#### 3.1 General Characterization

We now consider Bayesian Nash equilibria of the game which are characterized by (i) the reporting strategy of an agent as a function of her cost of lying  $l$  and the observed number  $y$  and (ii) the receiver’s beliefs about the likelihood that an agent lied which is a function of the reported number  $x$ . Denote the (mixed) strategy of an agent as

$$\zeta(l, y) = (p_{ly}^0, p_{ly}^1, \dots, p_{ly}^K),$$

which is a probability distribution over the  $K + 1$  pure reporting strategies, i.e., an agent with cost of lying  $l$  who has observed  $y$  reports number  $x$  with probability  $p_{ly}^x$ . The receiver’s belief that a report  $x$  was truthful is denoted as

$$r(x) = \Pr[y = x|x].$$

Then, an equilibrium is a set of mixed strategies and beliefs satisfying the following conditions:

$$\forall(l, x, y) : p_{ly}^x > 0 \text{ only if } x \in \arg \max_{x'} u(l, x', y), \quad (2)$$

$$\forall(l, y) : \sum_{x=0}^K p_{ly}^x = 1, \quad (3)$$

---

is the case. To be specific, the assumption states that there is a strictly positive probability that there are agents who never lie even if this would yield the highest possible lying payoff of  $K - l$ , while telling the truth would yield them the lowest possible truth-telling payoff of  $-\eta$ . Note that this assumption can also be viewed as a substitute for an equilibrium refinement as it guarantees that  $\Pr[y = x|x]$  is well defined for all  $x \in \mathcal{Y}$ .

$$\forall x : r(x) = \frac{\Pr [y = x \wedge x \text{ reported}]}{\Pr [x \text{ reported}]} = \frac{\int_0^\infty p_{lx}^x dF(l)}{\sum_{y=0}^K (\int_0^\infty p_{ly}^x dF(l))}. \quad (4)$$

The last condition ensures that beliefs are formed by Bayes rule.<sup>10,11</sup>

Denote by  $\mathcal{X}_L$  the set of numbers such that with a strictly positive probability they are reported by liars in equilibrium, i.e.,

$$\mathcal{X}_L = \{x \in \mathcal{Y} \mid r(x) < 1\}. \quad (5)$$

We first show that in any equilibrium  $\mathcal{X}_L$  is nonempty.

**Lemma 1** *In any equilibrium there is a strictly positive probability that the agent lies.*

**Proof.** Assume by contradiction that  $\Pr [y \neq x] = 0$ . Then,  $r(x) = 1$  for all  $x \in \mathcal{Y}$ , and there must be a positive measure of truth-telling agents for any interval of possible  $l$ . But then truth-telling agents with  $l < K - y$  observing  $y < K$  would have incentive to deviate and report  $x = K$ . Thus,  $\Pr [y \neq x] > 0$ . ■

When  $\mathcal{X}_L$  is a singleton, it is straightforward to show that the only feasible candidate equilibrium with this property is one in which  $\mathcal{X}_L = \{K\}$ . To see that, suppose that all agents who lie choose  $x' < K$ .<sup>12</sup> In that case,  $r(x') < r(K) = 1$ . But then the lying agents could earn a higher material payoff and reputation when reporting the highest possible report  $K$ .

Note, however, that such an equilibrium where  $\mathcal{X}_L = \{K\}$  may not exist. The reason is the following: suppose that sufficiently many liars report  $K$ . Then such a report can lead to a strong loss in reputation. In turn, when  $\eta$  is large enough liars will have an incentive to deviate and report a lower number. We can indeed show:

**Lemma 2** *In any equilibrium in which the set of reports chosen by liars  $\mathcal{X}_L$  is a singleton, all liars report  $x = K$ . If  $\eta$  is sufficiently large,  $\mathcal{X}_L$  can never be a singleton.*

<sup>10</sup>Recall that there are always agents who never lie (since  $F(K + \eta) < 1$ ). Hence any value of  $x$  is reported with strictly positive probability by a truthful agent on the equilibrium path.

<sup>11</sup>For our game, the defined equilibrium is equivalent to sequential equilibrium for psychological games developed by Battigalli and Dufwenberg (2009).

<sup>12</sup>More precisely, we mean by this "almost all agents who lie" (given that a deviation on a zero-measure of types does not affect  $\mathcal{X}_L$ ). The same note applies to further similar instances.

**Proof.** The first part of the lemma follows from the above considerations.

Let us show the second part. Assume that in equilibrium  $\mathcal{X}_L = \{K\}$ . Then, one can show that the total likelihood of lying  $\Pr[x = K|y \neq K]$  is bounded from below for any  $\eta$ . Indeed, consider agents observing  $y = 0$  with  $l < K - 1$ . These types will never tell the truth in equilibrium, since they have a strict incentive to deviate to  $x = K - 1$  earning  $K - 1 - l$ , which is then strictly higher than the payoff of 0 they earn from truth-telling (while there is no reputational loss in both cases). Thus, the probability of lying is bounded by the fraction of these types, i.e.,

$$\Pr[x = K|y \neq K] \geq \frac{1}{K+1}F(K-1). \quad (6)$$

Furthermore, note that agents observing  $K$  never report some  $x' \neq K$  (earning  $x' - l$ ) since otherwise it should hold  $x' > K - \eta(1 - r(K))$  so that all liars to  $K$  would have a strict incentive to also deviate to  $x'$  (being indifferent in terms of lying costs). Then, by Bayes rule

$$\begin{aligned} r(K) &= \frac{\Pr[y = K \wedge K \text{ reported}]}{\Pr[K \text{ reported}]} = \frac{\frac{1}{K+1}}{\frac{1}{K+1} + \Pr[x = K|y \neq K]} \quad (7) \\ &\leq \frac{\frac{1}{K+1}}{\frac{1}{K+1} + \frac{1}{K+1}F(K-1)} = \frac{1}{1 + F(K-1)}, \end{aligned}$$

i.e.,  $r(K)$  is bounded from above by  $\frac{1}{1+F(K-1)} < 1$ .

Moreover, if  $\mathcal{X}_L = \{K\}$  any agent who lies to  $K$  must prefer to report  $x = K$  to reporting  $x = K - 1$  or

$$K - 1 \leq K - \eta \cdot (1 - r(K)) \quad (8)$$

$$\Leftrightarrow 1 - \frac{1}{\eta} \leq r(K). \quad (9)$$

Conditions (7) and (9) thus yield

$$1 - \frac{1}{\eta} \leq \frac{1}{1 + F(K-1)}$$

which leads to a contradiction when  $\eta$  is sufficiently large. ■

Hence, when  $\eta$  is sufficiently large liars will not all report the same number in equilibrium, as otherwise someone reporting this number is perceived as a liar with a too high probability which makes a deviation attractive. However, there may be equilibria where liars report numbers from a larger set  $\mathcal{X}_L \subseteq \mathcal{Y}$ .

Intuitively, randomizing reports over a larger set makes it easier to disguise a lie. It is important to note that this can only be the case when liars are indifferent between all messages in  $\mathcal{X}_L$  or

$$x - \eta \cdot (1 - r(x)) = \rho \text{ for all } x \in \mathcal{X}_L \quad (10)$$

where  $\rho$  is some constant which is equal to the gross payoff from lying (without the lying costs  $l$ ). To see this, note that if for some  $x', x'' \in \mathcal{X}_L$  it would hold  $\rho(x') > \rho(x'')$ , then all liars to  $x''$  would have a strict incentive to deviate to  $x'$  given that their lying cost from reporting  $x'$  would either be the same as for  $x''$ , i.e.,  $l$ , or smaller (in case if  $y = x'$ ). We use this to further characterize necessary conditions for all feasible equilibria:

**Proposition 1** *In any equilibrium the following properties hold:*

(i) *The set of reports  $\mathcal{X}_L$  chosen by liars is characterized by a cut-off value  $x_L \in \{1, \dots, K\}$  such that  $\mathcal{X}_L = \{x \in \mathcal{Y} : x_L \leq x \leq K\}$ .*

(ii) *Agents who have observed a value  $y \geq x_L$  report their observed values truthfully.*

(iii) *The probability that someone who reports  $x$  is telling the truth is*

$$r(x) = \begin{cases} 1 - \frac{x-\rho}{\eta} & \text{if } x \geq x_L \\ 1 & \text{if } x < x_L, \end{cases} \quad (11)$$

where  $\rho$  is some constant in the interval  $(\max\{0, K - \eta\}, K)$ .<sup>13</sup>

(iv)  *$x_L$  is the smallest integer strictly larger than  $\rho$ :*

$$x_L = \min \{x \in \{1, \dots, K\} | x > \rho\}. \quad (12)$$

**Proof.** First, note that in any equilibrium  $\mathcal{X}_L$  is nonempty by Lemma 1. We now show that  $\mathcal{X}_L$  consists of the largest elements of  $\mathcal{Y}$  by contradiction: suppose that  $x_L$  is the smallest element of  $\mathcal{X}_L$  and there is a specific value  $x' > x_L$  which is not an element of  $\mathcal{X}_L$  so that  $r(x') = 1 > r(x_L)$ . Then  $u(x', y) > u(x_L, y)$  for all  $y \neq x_L$  (as both material and reputational payoffs are higher by reporting  $x'$ ) and liars to  $x_L$  would deviate and report  $x'$  instead. Thus,  $x_L \notin \mathcal{X}_L$  which is a contradiction.

Let us show that agents who observed  $y \in \mathcal{X}_L$  report their true value. Assume by contradiction that there exists  $y' \in \mathcal{X}_L$  such that some agent observing  $y'$  lies

<sup>13</sup>Note that if the set of possible numbers is bounded from below by  $\underline{y}$ , the interval of  $\rho$  changes to  $(\max\{\underline{y}, K - \eta\}, K)$ .

by reporting  $x = y''$ . For this agent, the following must hold

$$\begin{aligned} y' - \eta(1 - r(y')) &\leq y'' - \eta(1 - r(y'')) - l \\ \Rightarrow y' - \eta(1 - r(y')) &< y'' - \eta(1 - r(y'')). \end{aligned}$$

This implies that no agent observing  $y \neq y'$  would like to lie to  $y'$ , strictly preferring to report  $y''$ . Hence,  $y' \notin \mathcal{X}_L$  which is a contradiction. It follows that all untruthful reports are done by the agents observing  $y \notin \mathcal{X}_L$ .

The existence of the cut-off value, together with condition (10), leads to the characterization of the probability of truth-telling (11). Moreover,  $x_L \neq 0$  as the opposite implies that  $\mathcal{X}_L = \mathcal{Y}$  and hence nobody would lie since, as we have shown, agents observing  $y \in \mathcal{X}_L$  report their number truthfully. This is a contradiction by Lemma 1.

Since agents observing  $y \in \mathcal{X}_L$  tell the truth while liars must use all messages in  $\mathcal{X}_L$  by definition, we have

$$1 > r(x) > 0 \text{ for all } x \in \mathcal{X}_L, \quad (13)$$

which together with  $K \in \mathcal{X}_L$  and condition (10) leads to  $\rho \in (K - \eta, K)$ . Furthermore,  $\rho$  must be strictly positive (i.e., strictly higher than the lowest possible observation). Otherwise, we would have  $\rho - l < y$  for all  $(l, y)$  which would imply that liars observing  $y \notin \mathcal{X}_L$  have a strict incentive to deviate to truth-telling. Thus,  $\rho \in (\max\{0, K - \eta\}, K)$ .

Finally, we show that  $x_L$  must be equal to the smallest integer strictly larger than  $\rho$ . First, note that the equilibrium likelihood that a report  $x$  is truthful is

$$r(x) = 1 - \frac{x - \rho}{\eta} < 1 \text{ for all } x \geq x_L \quad (14)$$

by (10) and (13). This implies that  $x > \rho$  for all  $x \geq x_L$  and in turn we must have  $x_L > \rho$ . Assume now by contradiction that  $x_L - 1 > \rho$ . Then, any agent who lies would be better off reporting  $x_L - 1$  (getting at least  $x_L - 1 - l$ ) instead of an  $x \geq x_L$  (getting  $\rho - l$ ) which leads to a contradiction. Hence, we must have that

$$x_L - 1 \leq \rho < x_L. \quad (15)$$

■

In equilibrium we must thus observe a specific pattern in the association between a report and the reputation for truth-telling given this report. This

reputation is linearly decreasing within the set of reports chosen by liars: the loss in money when choosing a lower report must be exactly offset by an equivalent gain in reputation so that the sum of monetary and reputational payoffs remains constant at some level  $\rho$  (see (10)).

As Proposition 1 also shows, all agents who have observed a  $y \geq x_L$  tell the truth. The intuition for this is simple: reporting a different value within  $\mathcal{X}_L$  would lead to the identical sum of monetary and reputational payoffs but comes along with a utility loss due to the cost of lying  $l$ . A direct implication of this result is that there is no downward lying: agents who observed a number smaller than  $x_L$  either tell the truth or report a number larger or equal to  $x_L$ . Agents who observed a number larger or equal to  $x_L$  always tell the truth.

Finally, the gross payoff from lying  $\rho$  and the lower bound to the set of reports chosen by liars  $x_L$  are directly linked:  $x_L$  must be equal to the smallest integer strictly larger than  $\rho$ . The lying payoff  $\rho$  must be smaller than  $x_L$  as reporting  $x_L$  will always come with a loss in reputation. But it cannot be smaller than  $x_L - 1$  as otherwise liars would want to deviate and report  $x_L - 1$  avoiding the reputational loss.

We can now consider the behavior of agents who have observed a true value  $y$  strictly below the cut-off value  $x_L$  and thus either tell the truth or lie. When reporting the truth such an agent's utility is  $y$  and when lying she should obtain  $\rho - l$ . Hence, she will report truthfully if and only if<sup>14</sup>

$$\rho - y \leq l. \tag{16}$$

This implies:

**Lemma 3** *Consider an equilibrium with gross payoff from lying  $\rho$ . If an agent observes a value  $y < x_L$  she reports truthfully if and only if her cost of lying  $l$  is not smaller than a cut-off value  $l^*(y) = \rho - y$ .*

We can use Proposition 1 and Lemma 3 to characterize the equilibrium distribution of reported numbers:

**Proposition 2** *Consider an equilibrium with gross payoff from lying  $\rho$ . Then the probability that  $x$  is reported is*

$$\Pr[x \text{ reported}] = \begin{cases} \frac{\eta}{(K+1)(\eta-x+\rho)} & \text{if } x \geq x_L \\ \frac{1}{K+1} (1 - F(\rho - x)) & \text{if } x < x_L \end{cases} \tag{17}$$

---

<sup>14</sup>The specification of the strategy of the indifferent type with  $l = \rho - y$  does not matter for the subsequent results.

which is strictly increasing in  $x$ . Moreover,  $\Pr[x \text{ reported}] > 1/(K+1)$  for  $x \geq x_L$  and  $\Pr[x \text{ reported}] \leq 1/(K+1)$  for  $x < x_L$ .

**Proof.** By Bayes rule,

$$r(x) = \frac{\Pr[y = x \wedge x \text{ reported}]}{\Pr[x \text{ reported}]} \quad (18)$$

At the same time, by Proposition 1 for any  $x \geq x_L$

$$\Pr[y = x \wedge x \text{ reported}] = \frac{1}{K+1} \quad (19)$$

as  $x$  is observed with probability  $1/(K+1)$  and all agents who observe it tell the truth. We then obtain from (18) and (19) that

$$\Pr[x \text{ reported}] = \frac{1}{K+1} \frac{1}{r(x)} \text{ for } x \geq x_L, \quad (20)$$

which is strictly larger than  $1/(K+1)$  since  $r(x) < 1$  for any  $x \geq x_L$  by (5). By substituting  $r(x) = 1 - (x - \rho)/\eta$  from Proposition 1 we obtain that

$$\Pr[x \text{ reported}] = \frac{\eta}{(K+1)(\eta - x + \rho)} \quad (21)$$

for  $x \geq x_L$ .

For any  $x < x_L$  the likelihood that  $x$  is reported is (given Lemma 3)

$$\begin{aligned} \Pr[x \text{ reported}] &= \Pr[x \text{ observed}] \cdot \Pr[\rho - x \leq l] \\ &= \frac{1}{K+1} (1 - F(\rho - x)) \leq \frac{1}{K+1}. \end{aligned} \quad (22)$$

The claim that  $\Pr[x \text{ reported}]$  is strictly increasing in  $x$  for  $x \geq x_L$  follows from the fact that the right-hand side of (21) is strictly continuously increasing in  $x$  given that  $\rho > K - \eta$  by claim (iii) of Proposition 1. It is left to show that  $\Pr[x \text{ reported}]$  strictly increases in  $x$  also for  $x < x_L$ , which by (22) must be the case in case if  $\rho - x > 0$  for any  $x < x_L - 1$ . In turn, this condition is satisfied since  $\rho \geq x_L - 1$  by claim (iv) of Proposition 1. ■

Proposition 2 thus implies a specific pattern in the observed distribution of reports. First of all, all numbers above the threshold  $x_L$  are reported more frequently than they are actually observed, and all numbers below the threshold are

reported (weakly) less frequently than observed.<sup>15</sup> The reason is straightforward: the agent lies only when observing  $y < x_L$ , and she lies with a positive probability only to  $x \geq x_L$ .

Furthermore, the higher  $x$  the higher is the likelihood that  $x$  is reported. This is due to two effects: For  $x < x_L$  all agents who report such an  $x$  are telling the truth. But within this set the incentives to lie are stronger for lower observed numbers as here the material gain from lying is larger. For larger values of  $x$  ( $x \geq x_L$ ) the fact that the likelihood of a report  $x$  must increase in  $x$  is due to the reputational payoff obtained in equilibrium. As laid out in the above, all numbers reported by liars in equilibrium must lead to the same total payoff. In turn, any material gain must be offset by an equivalent reputational loss. As higher values of  $x$  come along with a higher material gain, more liars must thus choose to report them in equilibrium.

### 3.2 Existence and Uniqueness

In equilibrium, the total likelihood of sending a false message by an agent observing  $y < x_L$  must be equal to the total likelihood of receiving a false message. We now use this “accounting property” to show the existence and investigate the uniqueness of equilibrium.

**Theorem 1** *All equilibria of the game induce the same distribution of reported numbers characterized by  $\rho^* \in (\max\{0, K - \eta\}, K)$ , which is implicitly defined by the equation*

$$\sum_{y=0}^{x_L(\rho^*)-1} F(\rho^* - y) = \sum_{x=x_L(\rho^*)}^K \left( \frac{x - \rho^*}{\eta - x + \rho^*} \right) \quad (23)$$

*that always has a unique solution in this interval.*

**Proof.** By Proposition 1 and Lemma 3 the total fraction of liars is

$$\Pr[\text{Lie}] = \sum_{y=0}^{x_L(\rho)-1} \frac{1}{K+1} F(\rho - y) \quad (24)$$

as a function of  $\rho$ . At the same time we can determine the total probability that a reported number is a lie as

$$\Pr[\text{Lie}] = \sum_{x=x_L(\rho)}^K \Pr[x \text{ reported}] \cdot \Pr[x \text{ is a lie} | x \text{ reported}]. \quad (25)$$

---

<sup>15</sup>The number  $x_L - 1$  can still be reported with probability  $\frac{1}{K+1}$  if  $\rho = x_L - 1$ , which is not precluded by the incentive constraint (15).

Using Proposition 2 and the fact that  $\Pr[x \text{ is a lie} | x \text{ reported}] = 1 - r(x) = (x - \rho)/\eta$ , this is equal to

$$\sum_{x=x_L(\rho)}^K \frac{\eta}{(K+1)(\eta-x+\rho)} \frac{x-\rho}{\eta} = \sum_{x=x_L(\rho)}^K \frac{x-\rho}{(K+1)(\eta-x+\rho)}. \quad (26)$$

In equilibrium the fraction of liars (24) must be equal to the fraction of reported lies (26) which then leads to condition (23).

In a next step, we show that (23) always has a unique solution on  $(\max\{0, K - \eta\}, K)$ . In order to do so, consider the function

$$\theta(\rho) = \sum_{y=0}^{x_L(\rho)-1} F(\rho-y) - \sum_{x=x_L(\rho)}^K \frac{x-\rho}{\rho+\eta-x} \quad (27)$$

which is strictly increasing in  $\rho$  on  $(\max\{0, K - \eta\}, K)$ . Let us show that  $\theta(\rho) = 0$  always has a unique solution  $\rho^* \in (\max\{0, K - \eta\}, K)$ .

First, we can show that

$$\lim_{\rho \rightarrow \max\{0, K - \eta\}^+} \theta(\rho) < 0, \quad (28)$$

$$\lim_{\rho \rightarrow K^-} \theta(\rho) > 0. \quad (29)$$

Indeed, when  $\rho \rightarrow \max\{0, K - \eta\}$  (from above) then either  $\theta(\rho) \rightarrow -\infty$  (if  $\max\{0, K - \eta\} = K - \eta$ ) or  $\theta(\rho) \rightarrow -\sum_{x=1}^K x/(\eta-x) < 0$  (if  $\max\{0, K - \eta\} = 0$ ). When  $\rho \rightarrow K$  (from below) then  $x_L = K$  by claim (iv) of Proposition 1 so that  $\theta(\rho) \rightarrow \sum_{y=0}^{K-1} F(K-y) > 0$ .

We now show that  $\theta(\rho)$  is continuous on  $(\max\{0, K - \eta\}, K)$ . First note that  $\theta(\rho)$  is continuous when  $\rho$  is not an integer as in this case  $x_L(\rho)$  does not vary by claim (iv) of Proposition 1. We now demonstrate that the function is continuous also at integer values of  $\rho$ . Suppose we have an integer value of  $\rho' \in (\max\{0, K - \eta\}, K)$ . Then  $x_L(\rho') = \rho' + 1$  and

$$\theta(\rho') = \sum_{y=0}^{\rho'} F(\rho' - y) - \sum_{x=\rho'+1}^K \frac{x-\rho'}{\rho'+\eta-x}.$$

It is straightforward to see that  $\lim_{\varepsilon \rightarrow 0} \theta(\rho' + \varepsilon) = \theta(\rho')$  since  $x_L(\rho' + \varepsilon) = x_L(\rho')$  if  $\rho'$  is an integer and  $\varepsilon < 1$  by claim (iv) of Proposition 1. Now consider

$\lim_{\varepsilon \rightarrow 0} \theta(\rho' - \varepsilon)$ . First note that  $x_L(\rho' - \varepsilon) = \rho'$  for  $\varepsilon < 1$ . Hence,

$$\begin{aligned}
\theta(\rho' - \varepsilon) &= \sum_{y=0}^{\rho'-1} F(\rho' - \varepsilon - y) - \sum_{x=\rho'}^K \left( \frac{x - \rho' + \varepsilon}{\rho' - \varepsilon + \eta - x} \right) \\
&= \sum_{y=0}^{\rho'} F(\rho' - \varepsilon - y) - F(\rho' - \varepsilon - \rho') \\
&\quad - \sum_{x=\rho'+1}^K \left( \frac{x - \rho' + \varepsilon}{\rho' - \varepsilon + \eta - x} \right) - \left( \frac{\rho' - \rho' + \varepsilon}{\rho' - \varepsilon + \eta - \rho'} \right) \\
&= \sum_{y=0}^{\rho'} F(\rho' - \varepsilon - y) - F(-\varepsilon) \\
&\quad - \sum_{x=\rho'+1}^K \left( \frac{x - \rho' + \varepsilon}{\rho' - \varepsilon + \eta - x} \right) - \left( \frac{\varepsilon}{\eta - \varepsilon} \right).
\end{aligned}$$

But then, as  $\lim_{\varepsilon \rightarrow 0} F(-\varepsilon) = 0 = F(0)$ ,

$$\lim_{\varepsilon \rightarrow 0} \theta(\rho' - \varepsilon) = \sum_{y=0}^{\rho'} F(\rho' - y) - \sum_{x=\rho'+1}^K \left( \frac{x - \rho'}{\rho' + \eta - x} \right) = \theta(\rho'), \quad (30)$$

and thus the function is continuous. This together with (28), (29), the fact that  $\theta(\rho)$  is increasing on  $(\max\{0, K - \eta\}, K)$  and the intermediate value theorem implies that  $\theta(\rho) = 0$  has a unique solution  $\rho^*$  on  $(\max\{0, K - \eta\}, K)$ . Given that  $\rho$  uniquely pins down the equilibrium distribution of reported numbers by Proposition 1 (claim (iv)) and Proposition 2, the claim follows. ■

The intuition behind the condition (23) is the following. Recall that only agents who observed an  $y < x_L(\rho)$  lie. And only reports  $x \geq x_L(\rho)$  can be lies with a positive probability. In equilibrium the likelihood that an agent who observed  $y < x_L(\rho)$  decided to lie (the left-hand side of (23)) must be equal to the likelihood that an agent who reported  $x \geq x_L(\rho)$  actually lied (the right-hand side of (23)). The left-hand side of (23) is strictly increasing in  $\rho$ . Intuitively, the higher the gross payoff of lying  $\rho$  the more agents lie. At the same time, the right-hand side of (23) – the total probability that a reported number is a lie – is strictly decreasing in  $\rho$ . Intuitively, the larger the fraction of lies among the reported numbers the smaller is the reputation of somebody reporting an  $x \geq x_L$  and the smaller is thus  $\rho$ .

Therefore there is a unique value of  $\rho$  that solves this equation and, by Proposition 1 (claim (iv)) and Proposition 2, we thus can infer that any equilibrium must induce the same distribution of messages. We can now combine these insights to characterize the equilibrium strategies  $p_{ly}^x$  chosen by liars. Note that there are multiple equilibria but – as the following result shows – all are payoff equivalent to the unique mixed strategy equilibrium where all lying agents observing  $y < x_L$  pursue symmetric randomization strategies over  $\mathcal{X}_L$ , i.e.,  $p_{ly}^x = p^x$  for any  $y < x_L$ ,  $x \geq x_L$  and  $l < \rho - y$ .

**Proposition 3** *There is a unique mixed strategy equilibrium with symmetric lying strategies. In this equilibrium an agent lies if and only if  $y < x_L(\rho^*)$  and  $l < \rho^* - y$  and otherwise tells the truth. All liars report  $x \geq x_L(\rho^*)$  with probability*

$$p^x = \frac{1}{\sum_{y=0}^{x_L(\rho^*)-1} F(\rho^* - y)} \frac{x - \rho^*}{\eta - x + \rho^*}. \quad (31)$$

*All other equilibria yield the same distribution of messages and are payoff-equivalent.*

**Proof.** To derive  $p^x$  for given  $\rho$  recall that we know from Proposition 2 that in equilibrium

$$\Pr[x \text{ reported}] = \frac{\eta}{(K+1)(\eta - x + \rho)} \text{ for all } x \geq x_L.$$

But by the law of total probability this must be equal to

$$\begin{aligned} \Pr[x \text{ reported}] &= \frac{\eta}{(K+1)(\eta - x + \rho)} \\ &= \Pr[y = x \wedge x \text{ reported}] + \Pr[y \neq x \wedge x \text{ reported}] \\ &= \frac{1}{K+1} + \sum_{y=0}^{x_L-1} \Pr[x \text{ reported}|y] \frac{1}{K+1} \\ &= \frac{1}{K+1} + \sum_{y=0}^{x_L-1} F(\rho - y) p^x \frac{1}{K+1} \\ \Leftrightarrow \frac{\eta}{\eta - x + \rho} &= 1 + \sum_{y=0}^{x_L-1} F(\rho - y) p^x. \end{aligned}$$

Solving for  $p^x$  yields

$$p^x = \frac{1}{\sum_{y=0}^{x_L-1} F(\rho - y)} \frac{x - \rho}{\eta - x + \rho}.$$

Note that the reporting strategies characterized by  $p^x$ ,  $x_L$ ,  $l^*(y)$  and beliefs  $r(x)$  that are all uniquely pinned down by the equilibrium value of  $\rho^*$ , indeed constitute an equilibrium. In particular, the consistency of strategies with incentive constraint (2) and beliefs with Bayesian rule (4) has been shown previously. Besides, it is easy to verify that the remaining equilibrium condition (3) holds for the derived  $p^x$  since the condition (23) is satisfied. Hence, the existence of a unique (symmetric) equilibrium is guaranteed.

Finally, note that there can be multiple other equilibria yielding the same distribution of messages (uniquely characterized by  $\rho^*$  by Theorem 1). Such equilibria entail asymmetric lying strategies, i.e., strategies where  $p_{ly}^x$  is not identical for all  $y < x_L(\rho^*)$  and  $l < l^*(y)$ . However, since  $\rho^*$  uniquely defines  $r(x)$ ,  $x_L$  and  $l^*(y)$ , the total payoffs of all agents are the same in all these equilibria. ■

Proposition 3 implies that in the symmetric equilibrium the likelihood assigned by liars to report  $x$  is increasing and convex in  $x$ . Intuitively, once the reputation from reporting a number becomes low it gets less sensitive to a marginal increase in the probability mass of agents lying to this number (by Bayes' formula). Consequently, one requires a larger increase in  $p^x$  to generate the same amount of reputational loss.

At the same time, there are multiple payoff-equivalent equilibria yielding the same distribution of messages. This result is due to the fact that for a given cost of lying  $l$  and observed value  $y$  an agent is indifferent between all reports in  $x \geq x_L$  under the equilibrium distribution of reports. Hence, all strategies in which  $p_{ly}^x$  differs across  $l$  and  $y$  induce exactly the same payoffs as the equilibrium strategy where  $p_{ly}^x = p^x$  as far as the total distribution of reported numbers remains the same (equal to the unique equilibrium distribution pinned down by Theorem 1). In any asymmetric equilibrium the role of the random variable  $l$  (or  $y$ ) is thus purely that of a coordination device. In particular, there are also equilibria with partial lying where a given type of agent does not randomize between reports but instead plays a pure strategy, and the equilibrium distribution of reports arises due to different groups of lying agents playing different strategies.

We have thus shown that in any equilibrium agents with sufficiently small lying costs choose to lie when the realized number they observed is smaller than a cut-off value  $x_L$ . These agents then (generally) randomize among reports  $x \geq x_L$ .

### 3.3 Extension: Size-Dependent Cost of Lying

So far we assumed that the direct intrinsic costs of lying do not depend on the size of the lie. We now explore the robustness of our results when we assume that lying costs additionally include a linear size-dependent component. In particular, in this section we consider the extended utility function

$$u(l, x, y) = x - l \cdot I_{\mathcal{L}}(x, y) - a|x - y| - \eta \cdot \Pr[y \neq x|x], \quad (32)$$

where  $a \geq 0$  is a positive constant measuring sensitivity to the size of lying. We also assume that  $a < 1$ , since for  $a \geq 1$  the only possible equilibrium is complete truth-telling.<sup>16</sup> As in the above, we assume that  $F(0) = 0$ , and  $F(K + \eta) < 1$  which again assures that all numbers are reported with positive probability on the equilibrium path.

We can show that this game is outcome equivalent to a game with only fixed cost of lying (with adjusted parameters). To do this, it is useful to establish as a first step that there is still no downward lying:

**Proposition 4** *No agent reports a number lower than the observed one.*

**Proof.** Assume by contradiction that there is a type observing  $y'$  reporting a different number  $y'' < y'$  in equilibrium. Since she then loses both in terms of the material payoff and lying costs, she must obtain a reputational benefit from lying, i.e.,

$$r(y') < r(y''). \quad (33)$$

Moreover, reporting  $y''$  must yield a larger utility than truth-telling, i.e.,

$$y'' - \eta(1 - r(y'')) - a|y'' - y'| - l \geq y' - \eta(1 - r(y')) \quad (34)$$

$$\Rightarrow y'' - \eta(1 - r(y'')) - a|y'' - y'| > y' - \eta(1 - r(y')) \quad (35)$$

$$\Leftrightarrow (y'' - y') + \eta(r(y'') - r(y')) - a|y'' - y'| > 0. \quad (36)$$

We now show that this condition implies that any liar would then strictly prefer to report  $y''$  rather than  $y'$ . To see that consider a type who observes  $y \neq y', y''$ .

---

<sup>16</sup>Intuitively, in the latter case any deviation from one's own observation  $y$  to some  $x \neq y$  causes a loss in terms of material payoff net of lying costs. If the difference in the reputational payoffs  $r(x) - r(y)$  would offset this loss to still make this deviation profitable, then one can show that nobody would lie to  $y$  so that  $r(y) = 1$ , which contradicts  $r(x) - r(y) > 0$  (see the proof of Proposition 4 below for an analogous argument).

The difference in utility when reporting  $y''$  rather than  $y'$  is

$$\begin{aligned} & (y'' - y') + \eta(r(y'') - r(y')) - (a|y'' - y| - a|y' - y|) \\ \geq & (y'' - y') + \eta(r(y'') - r(y')) - a|y'' - y'| > 0, \end{aligned}$$

where the second inequality is by (36), and the first inequality follows from the easily verifiable arithmetic property that for any  $a, b, c \in \mathbb{R}$  we must have that

$$|a - c| - |b - c| \leq |a - b|.$$

This implies that no agent observing  $y \neq y'$  would like to lie to  $y'$ , strictly preferring to report  $y''$ . Hence,  $r(y') = 1$  which contradicts (33). ■

Since there is no downward lying in equilibrium, the agent's optimization problem can be replaced with maximizing the following expression over  $x \in \{y, \dots, K\}$ :

$$\max_{x \in \{y, \dots, K\}} (1 - a)x + ay - l \cdot I_{\mathcal{L}}(x, y) - \eta \cdot \Pr[y \neq x|x].$$

Since increasing utility by  $ay$  for any choice of  $x$  does not affect the optimal choice of  $x$  for a given  $y$ , this problem is equivalent to

$$\max_{x \in \{y, \dots, K\}} (1 - a)x - l \cdot I_{\mathcal{L}}(x, y) - \eta \cdot \Pr[y \neq x|x],$$

which is finally equivalent to

$$\max_{x \in \{y, \dots, K\}} x - \frac{l}{1 - a} \cdot I_{\mathcal{L}}(x, y) - \frac{\eta}{1 - a} \cdot \Pr[y \neq x|x].$$

Hence, the agent's choice problem is equivalent to the choice problem analyzed before – except that the parameters  $l$  and  $\eta$  are rescaled with a factor  $(1 - a)^{-1}$ .<sup>17</sup> Consequently, all previous results remain as before conditional on the rescaling of the parameters. In particular, since the distribution of the rescaled fixed lying costs is given by

$$\tilde{F}(z) = \Pr\left[\frac{l}{1 - a} \leq z\right] = \Pr[l \leq (1 - a)z] = F((1 - a)z),$$

---

<sup>17</sup>Note that in the previous analysis agents optimize over the set  $\{0, \dots, K\}$ . But since downward lying is never observed in equilibrium also in the previous model by Proposition 1 (claims (i) and (ii)), the restriction of the agent's action set to  $\{y, \dots, K\}$  does not affect the results.

the equilibrium condition in Theorem 1 becomes

$$\sum_{y=0}^{x_L(\rho^*)-1} F((1-a)(\rho^* - y)) = \sum_{x=x_L(\rho^*)}^K \left( \frac{x - \rho^*}{\tilde{\eta} - x + \rho^*} \right),$$

where  $\tilde{\eta} = \eta/(1-a)$  and  $\rho^* \in (\max\{0, K - \tilde{\eta}\}, K)$ . Correspondingly, the subsequent results derived in Section 4 for the baseline model are valid for the extended model with size-dependent lying costs conditional on the rescaling of the parameters.

Finally, we acknowledge that adding a nonlinear size-dependent lying cost does not generally allow to transform the agent's optimization problem into the original problem with only a fixed cost of lying. The equivalence hinges on the (piece-wise) linearity of the lying costs.

## 4 Comparative Statics

This section describes the comparative statics of the equilibrium with respect to a change in the image concerns, fixed lying costs and monetary stakes.

### 4.1 Effect of a Change in the Image Concerns

We now go back to our initial model and consider how a change in the agents' image concerns affects the equilibrium distribution of the reported numbers.

**Proposition 5** *If the agents' image concerns  $\eta$  increase, then*

- (i)  $x_L$  weakly decreases and
- (ii) the likelihood that an agent lies strictly decreases.

**Proof.** (i) We start with the accounting condition (23)

$$\theta(\rho, \eta) = \sum_{y=0}^{x_L(\rho)-1} F(\rho - y) - \sum_{x=x_L(\rho)}^K \frac{x - \rho}{\rho + \eta - x} = 0.$$

By Theorem 1, this implicitly defines a function  $\rho^*(\eta) \in (\max\{0, K - \eta\}, K)$  such that  $\theta(\rho^*(\eta), \eta) = 0$ . Consider some  $0 < \eta' < \eta''$ . As  $\theta(\rho^*(\eta'), \eta)$  is strictly increasing in  $\eta$  for  $\eta > \eta'$  (given that then  $\rho^*(\eta') > K - \eta' > K - \eta$ ) we obtain

$$\theta(\rho^*(\eta'), \eta'') > \theta(\rho^*(\eta'), \eta') = 0 = \theta(\rho^*(\eta''), \eta'') \quad (37)$$

so that

$$\theta(\rho^*(\eta'), \eta'') > \theta(\rho^*(\eta''), \eta''). \quad (38)$$

At the same time,  $\theta(\rho, \eta'')$  is strictly increasing in  $\rho$  on  $(\max\{0, K - \eta''\}, K)$ . This together with (38) and the fact that  $K > \rho^*(\eta') > \max\{0, K - \eta'\} \geq \max\{0, K - \eta''\}$  implies that

$$\rho^*(\eta') > \rho^*(\eta'') \quad (39)$$

for any  $\eta'' > \eta' > 0$ . It follows that  $\rho^*(\eta)$  is strictly decreasing in  $\eta$  and, by Proposition 1 (claim (iv)),  $x_L(\rho^*(\eta))$  must then be (weakly) decreasing in  $\eta$ .

(ii) Lemma 3 implies

$$\Pr[\text{Lie}] = \frac{1}{K+1} \sum_{y=0}^{x_L(\rho^*)-1} F(\rho^* - y). \quad (40)$$

From the proof of claim (i) we know that  $\rho^*$  is strictly decreasing in  $\eta$ , which together with  $\rho^* > 0$  by Theorem 1 leads to the claim. ■

Hence, stronger image concerns reduce the fraction of liars but enlarge the set of numbers they report. An intuition for the latter effect is that for high values of  $\eta$  liars suffer more from a loss in reputation. Hence, the incentives for liars to deviate to a lower number (which would not be conceived as a lie) become stronger. In equilibrium this pushes  $x_L$  downwards.

Next, we can show that in the limits of  $\eta$  the highest and the lowest feasible values of  $x_L$  are reached.

**Proposition 6** (i) *When image concerns become sufficiently large, all numbers besides 0 are reported by liars, i.e.,  $\lim_{\eta \rightarrow \infty} x_L(\eta) = 1$ .*

(ii) *When image concerns become sufficiently small, all agents who lie report  $K$ , i.e.,  $\lim_{\eta \rightarrow 0} x_L(\eta) = K$ .*

**Proof.** (i) Consider again condition (23):

$$\sum_{y=0}^{x_L(\rho^*(\eta))-1} F(\rho^*(\eta) - y) = \sum_{x=x_L(\rho^*(\eta))}^K \left( \frac{x - \rho^*(\eta)}{\rho^*(\eta) + \eta - x} \right).$$

Since  $\rho^*(\eta) > 0$  by Theorem 1, the right-hand side, which is continuously decreasing in  $\rho^*$  on  $(\max\{0, K - \eta\}, K)$ , is bounded from above for a given  $\eta > K$

by

$$\sum_{x=x_L(\rho^*(\eta))}^K \left( \frac{x}{\eta - x} \right)$$

which tends to 0 if  $\eta \rightarrow \infty$ . Consequently, the left-hand side must go to 0 as well.

Since

$$\sum_{y=0}^{x_L(\rho^*(\eta))-1} F(\rho^*(\eta) - y) \geq F(\rho^*(\eta) - 0) = F(\rho^*(\eta))$$

we then obtain that  $F(\rho^*(\eta)) \rightarrow 0$ . As  $F$  is a continuously increasing function and  $F(0) = 0$ , this implies that  $\lim_{\eta \rightarrow \infty} \rho^*(\eta) = 0$ . This together with Proposition 1 (claim (iv)) leads to the claim.

(ii) In equilibrium,  $\rho^* > K - \eta$  by Theorem 1. Consequently, at least for any  $\eta < 1$  we must have  $\rho^* > K - 1$  so that  $x_L(\eta) = K$  by Proposition 1 (claim (iv)).

■

Note that the result of Proposition 6 (i) has potentially interesting welfare implications. As we have seen already in the above, larger image concerns have a dual effect on reporting strategies: on the one hand, the fraction of lies is reduced, but on the other hand, it becomes harder to infer whether a person is telling the truth from observing a lower report as liars also report smaller numbers. When  $\eta$  is sufficiently large so that  $x_L = 1$ , an observer can never be sure that a reported message is not a lie (except for the lowest message). If such an observer would have to take a decision based on this information and mistakes are extremely costly, this might preclude her from taking a (potentially welfare-improving) action given the unavoidable risk of a mistake. Thus, image concerns of the agent may backfire for the receiver under certain circumstances. Appendix A provides a specific example of the receiver's preferences where this is indeed the case.<sup>18</sup> The extension shows formally that higher image concerns not to be perceived as a liar can backfire whenever it is important to either abstain from an action or be entirely sure to do the right thing. That is, such image concerns are harmful whenever even a small likelihood that a report is not true will render it useless.

---

<sup>18</sup>At the same time, one can show that the total probability of lying is converging to 0 if  $\eta \rightarrow \infty$  (the proof is available upon request). From this perspective, sufficiently high image concerns of the sender are beneficial to the receiver.

## 4.2 Effect of a Change in the Fixed Costs of Lying

Let us now consider how the equilibrium is affected by a change in the distribution of the fixed lying costs  $F$ . We consider an increase in  $F$  in the sense of first-order stochastic dominance (FOSD). That is, assume that the family of lying costs distributions can be parametrized by  $\lambda$  such that for any  $\lambda' < \lambda''$  and  $z > 0$

$$F_{\lambda'}(z) > F_{\lambda''}(z). \quad (41)$$

**Proposition 7** *If the distribution of lying costs  $F$  increases in the sense of FOSD, then*

- (i)  $x_L$  weakly increases and
- (ii) the likelihood that an agent lies strictly decreases.

**Proof.** (i) Consider the accounting condition (23) now denoted as a function of  $\rho$  and  $\lambda$ :

$$\theta(\rho, \lambda) = \sum_{y=0}^{x_L(\rho)-1} F_{\lambda}(\rho - y) - \sum_{x=x_L(\rho)}^K \frac{x - \rho}{\rho + \eta - x} = 0.$$

By Theorem 1, this implicitly defines a function  $\rho^*(\lambda) \in (\max\{0, K - \eta\}, K)$  such that  $\theta(\rho^*(\lambda), \lambda) = 0$ . By (41), for any given  $\rho > 0$  and any  $\lambda' < \lambda''$

$$\theta(\rho, \lambda') > \theta(\rho, \lambda''). \quad (42)$$

Hence, we must have that

$$\theta(\rho^*(\lambda''), \lambda') > \theta(\rho^*(\lambda''), \lambda'') = 0 = \theta(\rho^*(\lambda'), \lambda') \quad (43)$$

so that

$$\theta(\rho^*(\lambda''), \lambda') > \theta(\rho^*(\lambda'), \lambda'). \quad (44)$$

At the same time, for given  $\lambda$  and  $\eta$ ,  $\theta(\rho, \eta)$  is strictly increasing in  $\rho$  on  $(\max\{0, K - \eta\}, K)$ . Consequently, (44) implies  $\rho^*(\lambda'') > \rho^*(\lambda')$  for any any  $\lambda' < \lambda''$  and thus  $\rho^*(\lambda)$  must be strictly increasing in  $\lambda$ . By Proposition 1 (claim (iv)),  $x_L(\rho^*(\lambda))$  must then be (weakly) increasing in  $\lambda$ .

(ii) Lemma 3 and Theorem 1 imply

$$\Pr[\text{Lie}] = \frac{1}{K+1} \sum_{y=0}^{x_L(\rho^*)-1} F_{\lambda}(\rho^* - y) = \frac{1}{K+1} \sum_{x=x_L(\rho^*)}^K \frac{x - \rho^*}{\rho^* + \eta - x}. \quad (45)$$

From the proof of claim (i) we know that  $\rho^*$  is strictly increasing in  $\lambda$ , which together with the fact that the right-hand side of (45) is decreasing in  $\rho^*$  leads to the claim. ■

Thus, the effect of an increase in the fixed lying costs goes in the opposite direction of the effect of an increase in the image concerns not to be perceived as a liar: the range of reported lies *shrinks* as the fixed lying costs increase. The intuition for this effect is that at higher lying costs less agents lie and thus the reputational loss from reporting a high number decreases. In turn, the relative incentive to disguise lies by reporting smaller values is reduced.

Next, we consider again limit results. Let us take a convention that if  $\lambda \rightarrow \infty$  then  $F_\lambda(z) \rightarrow 0$ , and if  $\lambda \rightarrow 0$  then  $F_\lambda(z) \rightarrow 1$  for any  $z > 0$ . The next proposition shows that if lying costs get very large, agents lie only to the highest possible number. At the same time, it is not generally true that  $x_L$  converges to the minimal possible value of 1 if  $\lambda$  approaches 0 (given that it is increasing in  $\lambda$  by Proposition 7). The reason is that if  $\eta$  is sufficiently small, the incentives to disguise the lie by reporting lower numbers remain weak for any lying costs. In this case, agents may never report low numbers independently of the fixed lying costs.

**Proposition 8** (i) *When lying costs become sufficiently large, all agents who lie report  $K$ , i.e.,  $\lim_{\lambda \rightarrow \infty} x_L(\eta, \lambda) = K$ .*

(ii) *For every  $\tilde{x}_L \in \{1, \dots, K\}$  there exists an  $\eta' > 0$  such that  $\lim_{\lambda \rightarrow 0} x_L(\eta', \lambda) = \tilde{x}_L$ .*

**Proof.** (i) Consider the equilibrium condition (23) which implicitly defines  $\rho^*$  as a function of  $\lambda$ :

$$\sum_{y=0}^{x_L(\rho^*(\lambda))-1} F_\lambda(\rho^*(\lambda) - y) = \sum_{x=x_L(\rho^*(\lambda))}^K \left( \frac{x - \rho^*(\lambda)}{\rho^*(\lambda) + \eta - x} \right). \quad (46)$$

If  $\lambda \rightarrow \infty$ , then  $F_\lambda(\rho - y) \rightarrow 0$  for any given  $\rho$  and  $y$ . Since  $\rho^*(\lambda)$  is bounded from above by  $K$  (by Theorem 1), this implies that the left-hand side of (46) and hence the right-hand side of (46) converge to 0 for  $\lambda \rightarrow \infty$ . Since the denominator on the right-hand side of (46) is bounded from above and all terms in the corresponding sum are positive (given that  $x_L > \rho^*$  by Proposition 1 (claim (iv)) and  $\rho^* > K - \eta$  by Theorem 1), this is only possible if  $x_L = K$  and  $\rho^*(\lambda) \rightarrow K$ .

(ii) First, consider  $\tilde{x}_L = K$ . Then, the claim follows from the fact that for any  $\eta < 1$  we have that  $x_L(\eta) = K$  for any  $\lambda$  (see the proof of Proposition 6(ii)).

Next, consider any given  $\widetilde{x}_L \in \{1, \dots, K-1\}$ . Denote functions

$$\begin{aligned}
\theta(\rho, \lambda, \eta) &= \sum_{y=0}^{x_L(\rho)-1} F_\lambda(\rho - y) - \sum_{x=x_L(\rho)}^K \frac{x - \rho}{\rho + \eta - x}, \\
L_1(\eta) &= \lim_{\lambda \rightarrow 0} \theta(\widetilde{x}_L - 1, \lambda, \eta) \\
&= \lim_{\lambda \rightarrow 0} \left( \sum_{y=0}^{\widetilde{x}_L-1} F_\lambda(\widetilde{x}_L - 1 - y) - \sum_{x=\widetilde{x}_L}^K \frac{x - (\widetilde{x}_L - 1)}{\widetilde{x}_L - 1 + \eta - x} \right) \\
&= \widetilde{x}_L - 1 - \sum_{x=\widetilde{x}_L}^K \frac{x - (\widetilde{x}_L - 1)}{\widetilde{x}_L - 1 + \eta - x}, \\
L_2(\eta) &= \lim_{\lambda \rightarrow 0} \theta(\widetilde{x}_L, \lambda, \eta) = \lim_{\lambda \rightarrow 0} \left( \sum_{y=0}^{\widetilde{x}_L} F_\lambda(\widetilde{x}_L - y) - \sum_{x=\widetilde{x}_L+1}^K \frac{x - \widetilde{x}_L}{\widetilde{x}_L + \eta - x} \right) \\
&= \widetilde{x}_L - \sum_{x=\widetilde{x}_L+1}^K \frac{x - \widetilde{x}_L}{\widetilde{x}_L + \eta - x},
\end{aligned}$$

where we used the facts that  $F_\lambda(0) = 0$ , and that for any integer  $x$  we have that  $x_L(\rho)|_{\rho=x} = x + 1$ .

Let us show that there exists an  $\eta' \in (K - \widetilde{x}_L + 1, \infty)$  such that

$$L_1(\eta') < 0, \tag{47}$$

$$L_2(\eta') > 0. \tag{48}$$

Indeed, consider  $\eta \in (K - \widetilde{x}_L + 1, \infty)$ . In this domain, both  $L_1(\eta)$  and  $L_2(\eta)$  are continuously increasing in  $\eta$ , while  $L_2(\eta) > L_1(\eta)$ . Note also that  $L_1(\eta)$  has a unique root in this domain if  $\widetilde{x}_L > 1$  (since  $\lim_{\eta \rightarrow K - \widetilde{x}_L + 1} L_1(\eta) = -\infty$  and  $\lim_{\eta \rightarrow \infty} L_1(\eta) = \widetilde{x}_L - 1 > 0$ ). This implies that there exists an  $\eta'$  in a sufficient proximity to this root such that (47) and (48) hold. In the remaining case of  $\widetilde{x}_L = 1$  we have that  $L_1(\eta)$  is negative for all  $\eta > K - \widetilde{x}_L + 1$ , while  $L_2(\eta)$  is positive for sufficiently large  $\eta$ . Hence, (47) and (48) again hold for sufficiently large  $\eta$ .

Let us fix  $\eta'$  such that (47) and (48) hold. Given the definitions of  $L_1(\eta)$  and  $L_2(\eta)$ , these conditions imply that there exists a  $\underline{\lambda}$  such that for all  $\lambda \leq \underline{\lambda}$  we

have that

$$\begin{aligned}\theta(\widetilde{x}_L - 1, \lambda, \eta') &< 0, \\ \theta(\widetilde{x}_L, \lambda, \eta') &> 0.\end{aligned}$$

Note also that  $K > \widetilde{x}_L > \max\{0, K - \eta'\}$  (due to  $\eta' > K - \widetilde{x}_L + 1$  by construction). Consequently, for all  $\lambda \leq \underline{\lambda}$  the value of  $\rho$  yielding  $\theta(\rho, \lambda, \eta') = 0$  on the interval  $(\max\{0, K - \eta'\}, K)$  must lie between  $\widetilde{x}_L - 1$  and  $\widetilde{x}_L$  (since  $\theta$  is continuously increasing in  $\rho$  on this interval as shown in the proof of Theorem 1). Hence, by Theorem 1 for all  $\lambda \leq \underline{\lambda}$

$$\rho^*(\lambda, \eta') \in (\widetilde{x}_L - 1, \widetilde{x}_L). \quad (49)$$

This together with Proposition 1 (claim (iv)) implies that for  $\eta = \eta'$  we have  $x_L = \widetilde{x}_L$  for all  $\lambda \leq \underline{\lambda}$ . ■

Proposition 8 thus shows that partial lying to reports below  $K$  – the main feature of the characterized equilibrium – can emerge also if the fixed lying costs are negligible.<sup>19</sup> In fact, when the fixed lying costs become very small, the set of reports chosen by liars is not uniquely pinned down but is determined by the strength of the image concerns. Moreover, in this case we can partition all potential values of  $\eta > 0$  into disjoint intervals  $\mathcal{M}_1, \dots, \mathcal{M}_K$  (in a consecutive order) such that for every  $x \in \{1, \dots, K\}$  there is a corresponding interval of  $\eta$  from this partition which induces  $x_L = x$  when the costs of lying become sufficiently small, as shown in the next corollary.

**Corollary 1** *There is a unique partition of  $\mathbb{R}^+$  into  $K$  nonempty convex sets  $\mathcal{M}_1, \dots, \mathcal{M}_K$  such that  $x = \lim_{\lambda \rightarrow 0} x_L(\eta, \lambda)$  if and only if  $\eta \in \mathcal{M}_x$ . Moreover, for any  $\eta_i \in \mathcal{M}_i$  and  $\eta_j \in \mathcal{M}_j$  such that  $i < j$  we have that  $\eta_i > \eta_j$ .*

**Proof.** By Theorem 1 for each  $\eta$  and  $\lambda$  there exists a unique equilibrium value  $x_L(\eta, \lambda)$ . By Proposition 7(i)  $x_L(\eta, \lambda)$  is monotonically increasing in  $\lambda$ . Since it is also bounded, there is a unique value of  $x_L$  for each  $\eta$  to which  $x_L(\eta, \lambda)$  converges as  $\lambda \rightarrow 0$ . At the same time, by Proposition 8(ii) we know that for each  $x \in \{1, \dots, K\}$  there exists an  $\eta > 0$  such that  $x = \lim_{\lambda \rightarrow 0} x_L(\eta, \lambda)$ . Hence, there must be a partition of  $\mathbb{R}^+$  into  $K$  nonempty sets  $\mathcal{M}_1, \dots, \mathcal{M}_K$  such that  $x = \lim_{\lambda \rightarrow 0} x_L(\eta, \lambda)$  if and only if  $\eta \in \mathcal{M}_x$ .

<sup>19</sup>Note that we still need  $F(0) = 0$ , i.e., the assumption that people avoid the lie when they are otherwise indifferent between lying and telling the truth, to ensure the existence of the considered equilibrium (see the proof of Theorem 1).

Consider now any  $\eta_i \in \mathcal{M}_i$  and  $\eta_j \in \mathcal{M}_j$  such that  $i < j$ . By construction of sets  $\mathcal{M}$  we must have that  $x_L(\eta_i, \lambda) = i < j = x_L(\eta_j, \lambda)$  for sufficiently small values of  $\lambda$ . By Proposition 5 this implies  $\eta_i > \eta_j$ .

Finally, the monotonicity of the index  $i$  in  $\mathcal{M}_i$  with respect to  $\eta$  implies that all sets  $\mathcal{M}$  are convex. ■

The next result demonstrates that intrinsic lying costs are still required to explain the typical pattern observed in experiments that (i) at least several lowest numbers are not overreported (i.e.,  $x_L \geq 2$ ; that is there are at least two lowest reports which are reported with a probability between 0 and  $1/(K+1)$ ), and (ii) there is a significant share of subjects reporting even  $x = 0$ . In particular, if the fixed lying costs converge to 0 while  $\lim_{\lambda \rightarrow 0} x_L \geq 2$  (which happens if  $\eta$  is not too large by Corollary 1), the lowest number is almost never reported in contrast to this evidence.

**Proposition 9** *Fix any  $\eta \in \mathcal{M}_i$  such that  $i \geq 2$ . Then the equilibrium has the property that  $\lim_{\lambda \rightarrow 0} \Pr[x = 0] = 0$ .*

**Proof.** Note that for any  $\lambda$

$$\rho^*(\lambda) \geq x_L(\lambda) - 1 \geq 1, \tag{50}$$

where the first inequality is by Proposition 1 (claim (iv)), and the second inequality is due to Proposition 7 and  $\lim_{\lambda \rightarrow 0} x_L \geq 2$  (with the latter implied by  $i \geq 2$  and the definition of  $\mathcal{M}_i$ ). Hence,  $\rho^*(\lambda)$  is bounded from below by 1. By Lemma 3, the lying rate conditional on observing 0 is  $F_\lambda(\rho^*)$  which converges to 1 when  $\lambda \rightarrow 0$  (and  $\rho^*$  is bounded from below) such that  $\lim_{\lambda \rightarrow 0} \Pr[x = 0] = 0$ . ■

Hence, having nonnegligible fixed costs of lying in the model is necessary to explain the whole pattern of experimental data.<sup>20</sup>

### 4.3 Effect of a Change in the Monetary Stakes

As a final step in the theoretical analysis, we consider the effect of the size of monetary stakes on the structure of the equilibrium reporting behavior. Assume

---

<sup>20</sup>This is consistent with the results of Abeler, Nosenzo and Raymond (forthcoming) who came to a similar conclusion within their modeling framework.

now that the agent's utility is

$$u(l, x, y) = s \cdot x - l \cdot I_{\mathcal{L}}(x, y) - \eta \cdot \Pr[y \neq x|x],$$

where  $s$  denotes the stake size. Note that the agent's choice problem is equivalent to maximizing

$$\hat{u}(l, x, y) = x - \frac{l}{s} \cdot I_{\mathcal{L}}(x, y) - \frac{\eta}{s} \cdot \Pr[y \neq x|x] \quad (51)$$

such that an increase in stake size can be analogously expressed as a joint decrease in image concerns and lying costs. Intuitively, reputational concerns and fixed lying costs become relatively less important for the agent as the monetary reward for lying increases. Given (51), we can apply all of our earlier results substituting the image concerns parameter and costs of lying appropriately. In particular, an increase in the stake size (i.e.,  $s > 1$ ) corresponds to a shift in the distribution of the costs of lying in the sense of FOSD since for any  $z > 0$

$$F(z) = \Pr[l \leq z] < \Pr[l/s \leq z] = F_s(z), \quad (52)$$

where  $F_s(z)$  is the distribution of the normalized lying costs.

**Proposition 10** *An increase in monetary stakes raises the likelihood that an agent lies.*

**Proof.** For any  $\eta' < \eta''$  and  $\lambda' < \lambda''$  we must have that

$$\Pr[\text{Lie}|\eta', \lambda'] > \Pr[\text{Lie}|\eta', \lambda''] > \Pr[\text{Lie}|\eta'', \lambda''], \quad (53)$$

where the first inequality follows from Proposition 7 (ii) and the second from Proposition 5 (ii). Hence, any joint reduction in  $\eta$  and  $\lambda$  (which is outcome equivalent to an increase in  $s$  as shown above) causes a (strict) increase in the lying rate. ■

Proposition 10 is in line with the experimental evidence of Kajackaite and Gneezy (2017) who show that the total rate of lying significantly increases with the stake size once controlling for the perceived negative consequences from being eventually detected as a liar (which in turn might be correlated with monetary stakes). However, in some other studies not controlling for this (e.g., Mazar, Amir and Ariely, 2008) the evidence for the stake size effect on lying is ambiguous. The meta study by Abeler, Nosenzo and Raymond (forthcoming) find positive stake

size effects when analyzing only studies in which the stake size is varied within the same experiment. However, they don't find stake size effects when comparing lying behavior across different experiments in different laboratories.

At the same time, a plausible extension of our model would be consistent with observing only weak stake size effects. In particular, this could be explained by the possibility that a higher stake size leads to a higher marginal cost of lying. To see that consider our extension from Section 3.3 where we allowed costs of lying that linearly depend on the size of the lie. If we now assume that the marginal intrinsic cost of lying  $a$  is an increasing function of the stake size (i.e., of the marginal monetary benefit of lying by one report unit) the utility function becomes

$$u(l, x, y) = sx - l \cdot I_{\mathcal{L}}(x, y) - a(s) |x - y| - \eta \cdot \Pr[y \neq x|x].$$

We can apply the same considerations as in Section 3.3 to show that maximizing  $u(l, x, y)$  is equivalent to maximizing

$$\max_{x \in \{y, \dots, K\}} x - \frac{1}{s - a(s)} l \cdot I_{\mathcal{L}}(x, y) - \frac{1}{s - a(s)} \eta \cdot \Pr[y \neq x|x].$$

But note that

$$\frac{\partial}{\partial s} \left( \frac{1}{s - a(s)} \right) = - \frac{1 - a'(s)}{(s - a(s))^2}$$

which may have a negative sign depending on the shape of  $a(s)$ . In the latter case, the rate of lying should decrease by Propositions 5 and 7. Hence, when the intrinsic costs of lying depend on its monetary consequences, the effect of an increase in stake size  $s$  on the probability of lying can be attenuated and even may be (locally) negative.

With respect to the effect of the stake size on the *range* of reported lies (i.e.,  $x_L$ ) our model does not provide clear directional predictions. As outlined above, an increase in the stake size is equivalent to a joint decrease in both the reputational lying costs ( $\eta$ ) and the fixed lying costs ( $\lambda$ ). Yet, according to Propositions 5 and 7, these two effects push  $x_L$  in the opposite directions. The net effect depends on the parameter values, in particular on the form of distribution  $F$ . If the resulting decrease in the relative fixed lying costs is more prominent, the range of reported lies can expand with  $s$  (see Appendix B for a numerical example). Notably, this is in line with the effect of the stake size in Fischbacher and Föllmi-Heusi (2013) who observed a shift in reporting the second-highest number of 4 from 17.72 to 27.50 percent (with only the second

value being significantly different from 1/6) as the stake size was tripled.<sup>21</sup>

## 5 Empirical Calibration

The model predicts the main qualitative patterns observed in Fischbacher and Föllmi-Heusi (2013)-type cheating games. In particular, it rationalizes the existence of a threshold  $x_L$  with numbers below (above)  $x_L$  being under(over)reported (see Proposition 2). And, moreover, it shows why there are “partial lies”, i.e., why not all liars report the payoff maximizing number.

Of course, there may be other behavioral mechanisms at work in addition to those we capture in our model. But it is yet instructive to verify how well the model can be calibrated to fit the actual experimental data in Fischbacher and Föllmi-Heusi (2013). For the distribution of lying costs  $F$ , we assume that they are drawn from a normal distribution left-truncated at 0, where the underlying (nontruncated) distribution has mean 0 and standard deviation  $\sigma$ .<sup>22</sup> The calibration was obtained by finding the values of  $\eta$  and  $\sigma$  which minimize the mean squared error between the predicted and the observed values.<sup>23</sup>

The parameter values producing the best fit of the message frequencies predicted by the model to the data in the baseline treatment of Fischbacher and Föllmi-Heusi (2013) are  $\eta = 3.65$  and  $\sigma = 2.96$  yielding  $x_L = 4$  (Fig. 1). One can see that the model predicts the main data patterns from the experiment, in particular the substantial amount of partial lying to  $x = 4$ , as well as a positive fraction of subjects reporting 0.<sup>24</sup> Notably, setting  $\eta = 0$ , i.e., using the model without reputational concerns results in a much worse fit than that of the main model (see Fig. 2).<sup>25</sup>

The model is also capable to analyze specific treatment effects found in experimental cheating games. Gneezy, Kajackaite and Sobel (2018) compared two

---

<sup>21</sup>The rates of reporting smaller numbers were lower than 1/6 in both treatments. The control treatment was formed within the same subject sample, and hence was different from the baseline treatment reported in our Section 5.

<sup>22</sup>We assume that the distribution of lying costs is shifted towards 0 given the frequent experimental observation that the probability mass of individual social preference parameters is shifted towards payoff maximization (see, e.g., Engel, 2011).

<sup>23</sup>Note that using the extended model with size-dependent lying cost from Section 3.3 cannot improve the calibration since such model is structurally equivalent to the original model as shown in that section.

<sup>24</sup>Similarly good fit is obtained by calibrating the model to the aggregated data from the meta-study of Abeler, Nosenzo and Raymond (forthcoming).

<sup>25</sup>The mean squared error is 6.79 in the main model with image concerns and 29.90 in the benchmark model (in terms of percentage points). According to the F-test, the first model fits the data significantly better ( $p = 0.021$ ).

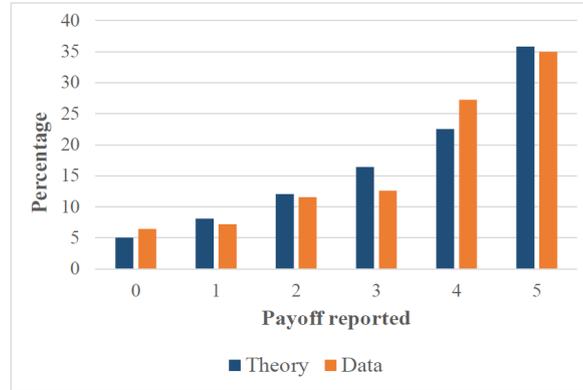


Figure 1: Empirical calibration of the main model to the baseline treatment in Fischbacher and Föllmi-Heusi (2013).

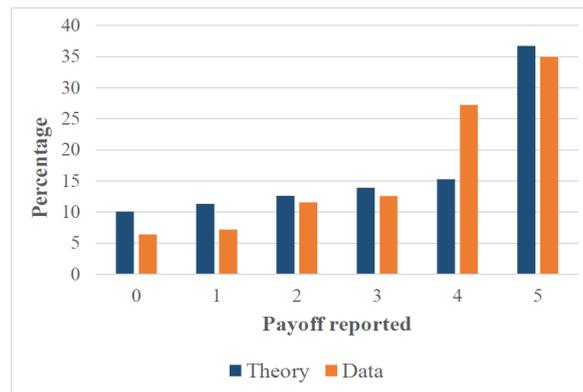


Figure 2: Empirical calibration of the model without reputational concerns ( $\eta = 0$ ) to the baseline treatment in Fischbacher and Föllmi-Heusi (2013).

treatments ("observed" and "non-observed") which differed in whether the experimenter was able to verify (ex-post) the true number rolled by the participant (since in the observed treatment the number was protocoled by the computer software).<sup>26</sup> Consider the observed treatment. Since subjects observed their true number privately on a computer screen (and later had to report it on a sheet of paper to the experimenter), they could still be uncertain whether the experimenter necessarily had verified their report by the time they received their payments (at which moment the image concerns may be considered to be most relevant). Hence, it is plausible to model the subject's utility function as

$$u(l, x, y) = x - l \cdot I_{\mathcal{L}}(x, y) - \eta \cdot (I_{\mathcal{L}}(x, y)\gamma + (1 - r(x))(1 - \gamma)),$$

where  $\gamma$  is the probability that the experimenter verifies the report.<sup>27</sup> This transforms into

$$u(l, x, y) = x - (l + \eta\gamma) \cdot I_{\mathcal{L}}(x, y) + \eta(1 - \gamma) \cdot (1 - r(x)).$$

Thus, when there is the possibility of verification the behavior of agents with parameter values  $\eta$  and  $l$  would be equivalent to the behavior of the hypothetical agents with adjusted parameters  $\tilde{\eta} = \eta(1 - \gamma)$  and  $\tilde{l} = l + \eta\gamma$  in our baseline setting. Consequently, the model would predict that in the observed treatment the *estimated* image concerns should decrease relative to the non-observed treatment (once the estimation is based on the baseline model in both cases). And analogously the *estimated* fixed lying costs should increase.

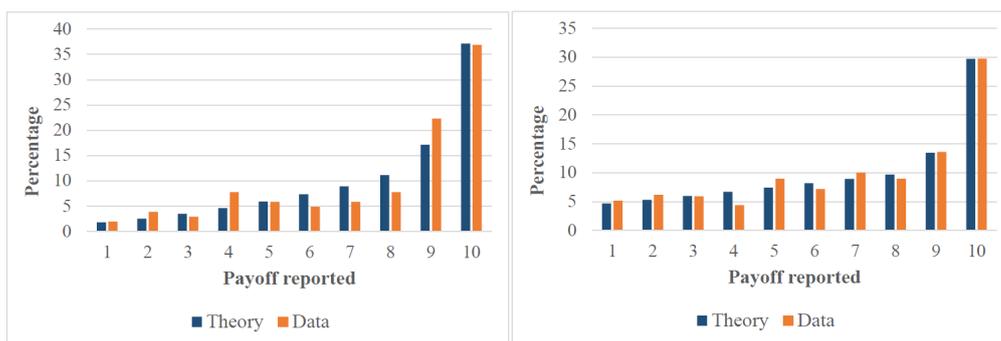
In line with this prediction, calibrating the empirical results from the two treatments in Gneezy, Kajackaite and Sobel (2018) with the baseline model results in that the observed treatment yields relatively lower estimate of image concerns ( $\eta = 2.45$  vs.  $\eta = 3.18$ ) and relatively higher estimate of the fixed costs of lying ( $\sigma = 10.18$  vs.  $\sigma = 4.98$ ), see Fig. 3.<sup>28</sup> Note also that the estimation of  $\eta$  in the non-observed treatment in Gneezy, Kajackaite and Sobel (2018) is close to that in the baseline treatment of Fischbacher and Föllmi-Heusi (2013) (3.18 vs. 3.65).

Finally, the data in Gneezy, Kajackaite and Sobel (2018) for the observed treatment allows to construct the distribution of reports conditional on a given number drawn by subjects (which was also known to the experimenter). As Fig.

<sup>26</sup>In particular, we consider their Numbers observed treatment and Basic non-observed treatment. In both of these treatments, the payoff was equal to the reported number.

<sup>27</sup>This captures that  $\Pr[Lie|x] = 1$  if the experimenter verifies the report while  $x \neq y$ , and  $\Pr[Lie|x] = 1 - r(x)$  if the experimenter does not verify the report.

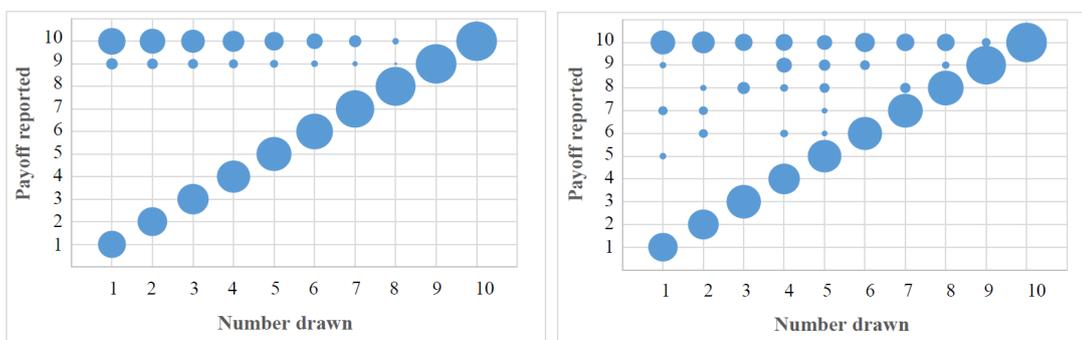
<sup>28</sup>The calibrated values of  $x_{\mathcal{L}}$  are 8 in the non-observed treatment and 9 in the observed treatment.



(a) Non-observed treatment

(a) Observed treatment

Figure 3: Empirical calibration of the main model to the experimental treatments in Gneezy, Kajackaite and Sobel (2018).



(a) Theory

(b) Data

Figure 4: Distribution of reports for each drawn number: a) in the calibrated model and b) in the observed treatment in Gneezy, Kajackaite and Sobel (2018).

4 shows, this empirical distribution is again close to the theoretical prediction (after substituting the previously calibrated parameter values).

## 6 Conclusion

We have shown that incomplete lying behavior naturally arises in the Fischbacher and Föllmi-Heusi (2013) setting when agents have (i) a fixed cost of lying and (ii) image concerns not to be perceived as a liar. When image concerns are sufficiently strong agents can “disguise” a lie only when liars randomize over a set of reports. This allows to reduce larger reputational losses occurring when all liars report the highest feasible report. In particular, as we have shown an agent’s reputation

to be honest must be strictly decreasing in the reported number so that any monetary gain is offset by an equivalent loss in reputation.

We furthermore have shown how the distribution of the reported numbers changes with the agents' image concerns. When they are weak, all liars report the number yielding the maximum feasible payoff. As image concerns increase, the lower bound of the set of numbers reported by the liars shifts to the left. In the limit, liars randomize over all but the lowest feasible number. Hence, higher image concerns of the agent may actually backfire for the receiver in terms of the usefulness of information: in the limit case, the receiver can never be sure that the reported number is not a lie, except for the lowest message. In contrast, higher fixed costs of lying always lead to a reduction in the range of reported lies.

The model can explain many stylized facts identified by Fischbacher and Föllmi-Heusi (2013) and the meta study by Abeler, Nosenzo and Raymond (forthcoming): a) the fraction of people reporting a payoff of 0 is positive, b) the fraction of people reporting 5 is above  $1/6$ ; c) the fraction of people reporting 4 is above  $1/6$ . Moreover, the model can replicate the observed difference in behavior once the experimenter is able to verify lying ex-post (as in Gneezy, Kajackaite and Sobel, 2018).

Overall, our model provides a useful workhorse for the analysis of empirical data from cheating experiments. The model may of course be extended in several directions. However, with only two components, i.e., fixed costs of lying and image concerns, our model attains already a reasonably good fit to the actually observed distribution of lies in cheating games. Future experiments may further verify the predictions of the model for instance by exogenously manipulating the strength of image concerns by varying the audience of outside observers (as in Ariely, Bracha and Meier, 2009), or by exogenously changing the costs of lying through imposition of random monitoring and punishments for liars.

## References

- Abeler, Johannes, Daniele Nosenzo, and Collin Raymond.** forthcoming. "Preferences for Truth-Telling." *Econometrica*.
- Ariely, Dan, Anat Bracha, and Stephan Meier.** 2009. "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially." *American Economic Review*, 99(1): 544–555.

- Battigalli, Pierpaolo, and Martin Dufwenberg.** 2009. “Dynamic Psychological Games.” *Journal of Economic Theory*, 144(1): 1–35.
- Bénabou, Roland, and Jean Tirole.** 2006. “Incentives and Prosocial Behavior.” *American Economic Review*, 96(5): 1652–1678.
- Bernheim, Bert D.** 1994. “A Theory of Conformity.” *Journal of Political Economy*, 102(5): 841–847.
- Brandts, Jordi, and Gary Charness.** 2003. “Truth or Consequences: An Experiment.” *Management Science*, 49(1): 116–130.
- Charness, Gary, and Martin Dufwenberg.** 2006. “Promises and Partnership.” *Econometrica*, 74(6): 1579–1601.
- Dufwenberg, Martin, Jr., and Martin Dufwenberg, Sr.** 2018. “Lies in Disguise—A Theoretical Analysis of Cheating.” *Journal of Economic Theory*, 175: 248–264.
- Ellingsen, Tore, and Magnus Johannesson.** 2008. “Pride and Prejudice: The Human Side of Incentive Theory.” *American Economic Review*, 98(3): 990–1008.
- Engel, Christoph.** 2011. “Dictator Games: A Meta Study.” *Experimental Economics*, 14(4): 583–610.
- Erat, Sanjiv, and Uri Gneezy.** 2012. “White Lies.” *Management Science*, 58(4): 723–733.
- Fischbacher, Urs, and Franziska Föllmi-Heusi.** 2013. “Lies in Disguise - An Experimental Study on Cheating.” *Journal of the European Economic Association*, 11(3): 525–547.
- Garbarino, Ellen, Robert Slonim, and Marie Claire Villeval.** 2016. “Loss Aversion and Lying Behavior: Theory, Estimation and Empirical Evidence.” GATE Working Paper Series 1631.
- Geanakoplos, John, David Pearce, and Ennio Stacchetti.** 1989. “Psychological Games and Sequential Rationality.” *Games and Economic Behavior*, 1(1): 60–79.
- Gneezy, Uri.** 2005. “Deception: The Role of Consequences.” *American Economic Review*, 95(1): 384–394.

- Gneezy, Uri, Agne Kajackaite, and Joel Sobel.** 2018. “Lying Aversion and the Size of the Lie.” *American Economic Review*, 108(2): 419–453.
- Gneezy, Uri, Bettina Rockenbach, and Marta Serra-Garcia.** 2013. “Measuring Lying Aversion.” *Journal of Economic Behavior & Organization*, 93: 293–300.
- Kajackaite, Agne, and Uri Gneezy.** 2017. “Incentives and cheating.” *Games and Economic Behavior*, 102: 433–444.
- Kartik, Navin.** 2009. “Strategic Communication with Lying Costs.” *Review of Economic Studies*, 76(4): 1359–1395.
- Kartik, Navin, Marco Ottaviani, and Francesco Squintani.** 2007. “Credulity, Lies, and Costly Talk.” *Journal of Economic Theory*, 134(1): 93–116.
- Lundquist, Tobias, Tore Ellingsen, Erik Gribbe, and Magnus Johannesson.** 2009. “The Aversion to Lying.” *Journal of Economic Behavior & Organization*, 70(1-2): 81–92.
- Mazar, Nina, On Amir, and Dan Ariely.** 2008. “The Dishonesty of Honest People: A Theory of Self-Concept Maintenance.” *Journal of Marketing Research*, 45(6): 633–644.
- Ottaviani, Marco, and Peter Norman Sørensen.** 2006a. “Reputational Cheap Talk.” *RAND Journal of Economics*, 37(1): 155–175.
- Ottaviani, Marco, and Peter Norman Sørensen.** 2006b. “The Strategy of Professional Forecasting.” *Journal of Financial Economics*, 81(2): 441–466.
- Sutter, Matthias.** 2009. “Deception Through Telling the Truth?! Experimental Evidence From Individuals and Teams.” *Economic Journal*, 119(534): 47–60.

## **Appendix A: Adverse Effect of Reputational Concerns**

In this section, we consider an example where an increase in the agent’s reputational concerns  $\eta$  may backfire for a receiver of the agent’s message. Assume that the receiver is a potential buyer of a credence good (say different types of surgery,

or different types of investment products), who can choose a product out of  $K + 1$  options from 0 to  $K$ . He also has an outside option of abstainment from purchase. Thus, his action space is  $\{0, 1, \dots, K\} \cup A$ , where a number denotes buying a specific product variety, while  $A$  denotes abstainment. Before the game, Nature randomly chooses one option to be profitable with each option being chosen with probability  $1/(K + 1)$ . The profitable option is denoted as  $y \in \{0, 1, \dots, K\}$ . The buyer's payoff is

$$\pi_R = \begin{cases} 0 & \text{if } a = A, \\ y & \text{if } a = y, \\ -L & \text{if } a \notin \{y, A\}. \end{cases}$$

Thus, the buyer obtains a positive payoff if and only if he buys the profitable option (larger than zero), while different options yield different potential profits. We also assume that  $L$  is sufficiently large, i.e., the buyer suffers a severe loss in case he buys a nonprofitable option.

Then, assume that the buyer observes a report  $x$  made by an agent prior to buying. The agent observes  $y$  and has preferences as specified in our model such that she benefits from reporting higher numbers but her payoff does not depend on the buyer's choice (for instance, the agent is an "expert" with reputational concerns who does not sell the product but has a preference for giving specific recommendations beyond its true value for the buyer).

In this case, if  $L$  is sufficiently large, the buyer would avoid buying the product if there is a positive likelihood that the agent has lied, i.e., the buyer's optimal strategy would be (assuming without loss of generality that the buyer prefers buying over  $A$  in case of indifference):

$$a^* = \begin{cases} A & \text{if } x \geq x_L \\ y & \text{if } x < x_L, \end{cases}$$

where  $x_L$  is as defined in our model. Consequently, the buyer's ex-ante payoff is

$$E[\pi_R] = \sum_{z=0}^{x_L-1} (\text{Pr}[x = z] \cdot z).$$

This implies that  $E[\pi_R] > 0$  if and only if  $x_L > 1$ . Hence, by Propositions 5 and 6, the buyer is better off when image concerns  $\eta$  are not too big (such that  $x_L > 1$ ) as for large enough  $\eta$  we have that  $x_L = 1$ .

In other words, too strong reputational concerns make it attractive to lie even towards very low reports that are unattractive for the agent. In turn, even after

observing an unattractive recommendation in this case (as long as it is not the most unattractive one), the buyer cannot be sure that such a report comes from an honest agent, in which case he prefers to abstain from decision. Hence, the advice becomes effectively useless. If, however, agents have weak reputational concerns, they would never report unattractive options – thus any such advice must come from honest agents, and hence can be followed without risk.

## Appendix B: Numerical Simulation of the Effect of the Monetary Stakes on $x_L$

When we substitute  $l$  and  $\eta$  according to a shift in the stake size to  $s \cdot x$ , equilibrium condition (23) becomes

$$\theta(\rho, s) = \sum_{y=0}^{x_L(\rho, s)-1} F_s(\rho - y) - \sum_{x=x_L(\rho, s)}^K \frac{x - \rho}{\rho + \frac{\eta}{s} - x} = 0, \quad (54)$$

where  $F_s(z)$  is given by (52) (see Section 4.3). The equilibrium is characterized by a unique value  $\rho^* \in (\max\{0, K - \eta/s\}, K)$  solving the equation.

Consider an example with  $K = 5$ ,  $\eta = 3.65$ , and  $F$  being a normal distribution left-truncated at 0, where the underlying (nontruncated) distribution has mean  $\mu$  and standard deviation  $\sigma = 1$ , so that  $F_s(z) = F(z|\mu/s, \sigma/s)$ . Then, if  $\mu = 0$  and  $s$  increases from 1 to 2, the value of  $\rho^* \in (\max\{0, K - \eta/s\}, K)$  solving equation (54) changes from 2.681 to 3.584, and hence  $x_L$  increases from 3 to 4. At the same time, if  $\mu = 5$ , then the same increase in  $s$  leads to a decrease in  $\rho^*$  from 4.237 to 3.840 so that  $x_L$  drops from 5 to 4.