Impact of lying aversion and prosociality on cheating[*]

Daniel Parra[†].

[Click here for the latest version of the paper]

Abstract

When individuals decide whether or not to lie, they compare the monetary benefits with the psychological cost of violating their norms. In addition, individuals are more likely to lie when their lies benefit others. This paper compares the impact of the aversion to lying and prosociality on cheating. I first present a model that incorporates heterogeneous lying costs and prosociality as a part of the individual's preferences. I show that individuals are mostly honest when some else has lied on their behalf. At the same time, if lying generates a positive externality, individuals lie more due to prosocial motives. I test these predictions in two online experiments. I show that participants are more dishonest when their lies benefit others. More importantly, I present evidence that, on average, the prosocial motive is stronger than the lying aversion motive. Further results show that individuals care about their influence on others' outcomes rather than taking actions that signal a prosocial intention but do not impact others' outcomes.

JEL Codes: C91, D02, D90.

Keywords: Cheating; Dishonesty; Prosociality; Psychological lying costs.

This version: October 27, 2021

---

[*] I am grateful to Kai Barron, Tilman Fries, Jeanne Hagenbach, Agne Kajackaite, Johannes Leutgeb, Cesar Mantilla, Robert Stüber, Yuliet Verbel, and audiences at WZB, BEBES, the 2021 ESA Global Online Meetings for helpful comments. This research used generic funds provided by the WZB Berlin Social Science Center. The usual disclaimer applies.

[†] WZB Berlin Social Science Center and Berlin School of Economics. E-mail: daniel.parra@wzb.eu

1 Introduction

Evidence on lying has shown that, even if there is no punishment for lying and there are personal benefits from doing so, people lie only moderately (Abeler et al., 2019). People's aversion to lying can explain the moderate extent of dishonesty. Theoretical models usually represent this lying aversion as psychological costs of lying.[1] These costs capture the idea that some individuals do not lie because they dislike violating their internal moral norm of being honest or because they want to appear honest. However, psychological lying costs might be reduced when one can lie to benefit oneself as well as others. This effect occurs because people might use prosociality to make lying easier. The main issue is that having more dishonesty is detrimental to society. For instance, in his classical model, Akerlof (1970) highlights the central role of dishonesty in markets with asymmetric information. He argues that, in the presence of information asymmetries, dishonesty can lead to market failures. In particular, the social damage generated by dishonesty includes the direct cost to the deceived individual and other indirect costs, such as eroding the incentives to produce high-quality goods.

In general, when one's lie generates a positive externality, two effects go in opposite directions. Lying aversion makes telling a lie costly, but the prosocial lie generates some utility. A vast literature in economics shows that prosocial incentives are important motivators in people's decisions (Andreoni, 1990; Andreoni and Miller, 2002; Charness and Rabin, 2002; Bénabou and Tirole, 2006; Ariely et al., 2009; DellaVigna et al., 2012). To illustrate how lying aversion and prosociality interact, imagine an intermediary selling a used car. The intermediary has incentives to lie about the actual quality of the car. They earn a higher commission if they sell the car for a higher price but incur psychological lying costs if they lie. However, all else equal, they may also feel less unethical by lying about the car's quality because they benefit the owner. A sales representative faces a similar trade-off between prosociality and lying aversion when they can lie to get a team bonus the CEO has promised after the sales team reaches a certain threshold. This duality is present in the political sphere too. For instance, some high-level politicians use a Chief of Staff (CoS) to have someone protecting their political interests. The CoS has to make decisions involving covering wrong actions of the politician, but given that they do it to benefit their politician, their prosocial lies might reduce their costs of lying.

This paper compares the impact of lying aversion and prosociality on cheating. To do so, I study lying in a dyadic setting. If at least one member of the dyad lies in the game, both benefit from the lie, but there are no additional gains if both lie. I use a dyadic game because it creates a situation where people can avoid lying, relying on others' incentives, or they can tell a prosocial lie. Hence, in this strategic situation, there is a trade-off. On the one hand, lying aversion implies that people are primarily honest when others are likely to lie on their behalf. On the other hand, prosociality implies that people are prone to lie when they benefit others. The paper aims to disentangle these two motives and assess which one is a stronger motivator.

---

[1] Psychological lying costs include the intrinsic costs of lying and the image concerns.

I first present a theoretical framework that incorporates heterogeneous psychological lying costs and prosociality in individual preferences. I include prosociality as a parameter that reduces the lying costs. Then, I use experimental data to assess the model's predictions empirically. I have run two online experiments. The first experiment used a cheating game where participants drew a Black card or an Orange card and then reported their color. The experimenter observed the random draw and the report. In this experiment, lying was very low, which made it difficult to find any treatment difference. To provide more privacy to participants, I ran a second experiment where the random draw was in participants' minds. In both experiments, individuals draw a low-paying state or a high-paying state randomly. The probabilities of each state are 0.8 for the low-paying and 0.2 for the high-paying. The random draw is known by the individual but not by their partner. They are asked to report what they draw to determine their payoffs. That is, the random draw is not relevant for the monetary payoffs but just the individuals' reports.

The two experiments had the same four treatments. The experiments were pre-registered in AEA RCT Registry.[2] In the first treatment called AVOID, two players report the result of the private random draw sequentially. Both get a higher monetary payoff if at least one individual reports the high-paying state. Therefore, the first mover can avoid lying if lying aversion is a stronger motivator than prosociality. In a second treatment called NO AVOID, the first player reports their random draw. However, the second player can no longer report, but a computer program reports the random draw truthfully. That is, by design, in NO AVOID it is common knowledge that the second player's report will be truthful. This variation makes it more difficult to avoid the lying costs in NO AVOID than in AVOID. As a next step, to eliminate the positive externality, I include a third treatment called NO EXTERNALITY. In this treatment, the first mover report does not benefit the second mover, allowing me to disentangle the prosocial motive from the lying aversion motive. Finally, I use a fourth treatment called SIMULTANEOUS where both players play at the same time. In this last treatment, there is a trade-off between strategically reporting the high-paying state to earn more and losing the prosocial motive because both reported the high-paying state. The prosocial motive is lost because participants hold a consequentialist view of prosociality. In other words, consequentialist prosocial lying implies that lying makes people feel better when the lie has an actual positive consequence on others.

---

[2] The registration numbers are AEARCTR-0006881 and AEARCTR-0007214.

The results of the second experiment show that the second mover in AVOID lies less when the first-mover reported the high-paying outcome, which is consistent with lying aversion. Surprisingly, I did not find any difference in the lying rates of the first-mover between AVOID and NO AVOID suggesting that prosociality was a strong driver of lying behavior. However, I found that first-movers lie more in NO AVOID than in NO EXTERNALITY indicating that people lie more when they benefit others which is in line with Wiltermuth (2011), Gino et al. (2013), and Levine and Schweitzer (2015). Furthermore, combining these results, I show that prosociality outweighs lying aversion. The final result of the experiment is that lying is higher in AVOID than in SIMULTANEOUS. This result confirms that individuals care about the actual benefit they generate in others rather than the prosocial intention of their actions.

This paper contributes to the fast-growing literature on lying behavior. This body of literature has argued that the deviation from the world full of liars that economic theory assumes can be explained by people's disutility when they are dishonest. Kajackaite and Gneezy (2017) show that individuals follow a cost-benefit analysis in which they evaluate the psychological cost of lying and the incentives to lie. Gneezy et al. (2018), Abeler et al. (2019), and Khalmetski and Sliwka (2019) present evidence that individuals indeed have psychological costs of lying that can be divided into intrinsic costs of lying and reputation. Dufwenberg and Dufwenberg (2018) also explain lying behavior by the cost of lying. However, they argue that this cost increases proportionally to the amount in which the individual is perceived to cheat, making the lying costs extrinsic. I contribute to this literature by showing that individuals do avoid lying when someone lies on their behalf. However, they do not avoid lying when their lies benefit others. Hence, I show that even if the psychological costs of lying make people lie less, prosocial lying is a stronger driver of behavior.

This paper closely relates to the literature on collaborative lying (for a survey see Leib et al., 2021). These studies use games in which participants play in groups, and all of them need to lie to increase their earnings (e.g. Conrads et al., 2013; Weisel and Shalvi, 2015; Muehlheusser et al., 2015; Kocher et al., 2018; Rilke et al., 2021). In other words, in this body of research, lies are strategic complements. Hence, collaborative lying research focuses on situations where coordination in dishonesty is at the center; thus, it is impossible to rely on others to save on lying costs. Although I also use a group setting, I study a different situation where dishonesty is not complementary but a substitute. Therefore, individuals can avoid being dishonest by relying on others. I show that individuals use prosociality to justify lying even in situations where collaboration is unnecessary to increase payoffs.

This paper also relates to the studies which analyze the impact of positive externalities on lying behavior. Wiltermuth (2011), Gino et al. (2013), and Levine and Schweitzer (2015) show that people are more likely to lie if they benefit others. Levine and Schweitzer (2015) show that prosocial lying enhances trust in group settings, which may explain why people are willing to lie for others. I add to this body of evidence by showing that prosociality is strong enough to outweigh lying aversion. Although, the effect of prosociality on dishonesty vanishes when the actual impact on others' payoffs is uncertain. This result points to a consequentialist view of prosocial lies. To put it another way, individuals' utility depends on the actual consequences of their actions on others rather than on the intention of benefiting them.

The paper proceeds as follows. Section 2 presents a theoretical model of lying with heterogeneous psychological lying costs and prosociality. It also presents four treatments that disentangle lying aversion and prosociality with their respective hypotheses. Section 3 presents evidence from two experimental studies that test the hypotheses of the model. Section 4 discusses the findings from the experiments and interprets them using the benchmark presented in Section 2. Section 5 concludes.

2 Theoretical framework, experimental design, and hypotheses

2.1 Individual's preferences

The lying models presented by Dufwenberg and Dufwenberg (2018), Gneezy et al. (2018), Abeler et al. (2019), and Khalmetski and Sliwka (2019) are fundamental to understand why people lie in situations without externalities.[3] I use them as a starting point and add strategic interaction to study the willingness to avoid the lying costs and lie prosocially. Specifically, I study situations where individuals interact in dyads. I denote the members of each dyad as P_i where $i \in \{1, 2\}$.

Players play a binary lying game. That is, each player draws a state $x_i \in \mathcal{X} = \{0, 1\}$. The probability of $x_i = 0$ is 0.8 and the probability of $x_i = 1$ is 0.2. I use these probabilities because they will generate more players drawing 0 than in a typical coin toss; therefore, more individuals will face the situation where they can lie to improve their payoffs. Players send a report $r_i \in \mathcal{X}$. The players' payoffs are interdependent. If at least one player, P_1 or P_2, reports 1, each of them will get a monetary payoff v_h. If both report 0, they will get v_l. To ease the notation, I normalize v_l to zero and v_h to 1. In this context, lies are under the category of Pareto White Lies (Erat and Gneezy, 2011) because they help others and benefit the liar.

---

[3] These models make use of psychological game theory to model behavior. They use the experimenter as an observer that affects the individual utility. Hence, they study strategic games with one player making decisions and a third party who does not take any particular action.

Individuals' preferences depend on three elements that determine the willingness to lie or tell the truth. First, they get utility from the monetary payoff $v_i \in \{v_h, v_l\}$ that depends on their report. All else equal, they have extrinsic incentives to report 1 regardless of their actual random draw x_i. Second, individuals dislike lying. Lying aversion is represented formally by some psychological costs (c_i) that they incur when they misreport their random draw (Gneezy et al., 2018; Abeler et al., 2019; Khalmetski and Sliwka, 2019). The psychological lying cost include the intrinsic costs of lying and the image costs. These costs, represented by c_i, are distributed among the population according to $c_i \sim U[0, \bar{c}]$. Hence, the cumulative density function of c_i is $F(c_i) = \frac{c_i}{\bar{c}}$. The heterogeneity in the psychological lying costs considers that some people are more morally inclined than others.

Third, I formally include prosociality in the utility function inspired by the insights provided by Wiltermuth (2011), Gino et al. (2013), and Levine and Schweitzer (2015). In particular, individuals get some satisfaction (θ) when they benefit others with their report, i.e. when they generate a positive externality. In the case of θ, I impose $0 \leq \theta \leq 1$ so that the prosocial lying utility is non-negative but never higher than the utility of the own monetary reward. The standard models omit prosociality in the decision on whether to lie or not. In my model, when an agent lies and benefits others, the prosocial lie reduces individuals' psychological lying cost. Moreover, I assume that $c_i(r_i = x_i) = 0$, and $1 + \theta < \bar{c}$. The last condition is used to rule out the uninteresting case where all individuals have a psychological cost of lying so small that everyone lies. With this assumption, the individual with the highest lying cost will always tell the truth.

The remaining question at this point is: does the utility derived from prosocial lies depend only on actions or also on consequences? Some models use warm-glow and altruism to explain giving (Andreoni, 1990) which implies that people care about their intentions to give. Conversely, I assume that individuals' utility depends on the outcome of their actions. Therefore, the positive externality reduces the cost of lying only if the marginal benefit of one's report on the partner is 1. In other words, the utility from prosociality (θ) when their partner reports 1 is 0 regardless of one's report. [4] Arguably, it is more difficult to justify a dishonest behavior with prosociality when in the absence of one's report, the payoff of the other would be the same. With these three elements, I represent individuals' preferences by the following function:

$$U_i(x_i, r_i, r_j) = r_i + r_j - r_i r_j - 1_{x_i \neq r_i}(c_i + \theta r_i(1 - r_j)) \tag{1}$$

---

[4] The alternative way to incorporate the positive externality's impact would be to assume that only by lying and reporting 1 they feel good. This view would represent deontological prosocial lying where the intention of benefiting others matters regardless of the actual consequence.

2.2 Treatments

2.2.1 Treatment 1: AVOID

In the main treatment, AVOID, I study a two stage lying game where players' lies are substitutes. P_1 draws $x_1 \in \mathcal{X}$ and sends a report $r_1 \in \mathcal{X}$ to P_2. After learning r_1, P_2 draws $x_2 \in \mathcal{X}$ and sends a report $r_2 \in \mathcal{X}$. Note that x_i is only known by P_i, but not by the other player.

I use backward induction to analyze the strategic context of the game. When P_1 reports 1, P_2 has no strict incentives to lie. Whereas, when P_1 reports 0, P_2's best response is 1 if $1 + \theta > c_i$. That is, if the second-mover considers that the combination of the monetary incentives and the satisfaction of benefiting others exceed the costs of lying, they will report 1 regardless of x_2. Importantly, in this game there is no downward lying in equilibrium. If individuals draw 1 and report 0 they incur the cost of lying without getting the monetary payoffs or the benefit of the positive externality. So, in equilibrium individuals only lie if they draw $x_i = 0$ by reporting $r_i = 1$.

Let $\hat{c_i}$ be the lying cost threshold where individuals are indifferent between lying or not. This threshold for P_2, when P_1 reports 0, is $\hat{c_2}(r_1 = 0) = 1+\theta$. Hence, the probability that P_2 lies after P_1 reports 0 is the expected proportion of players with $\hat{c_2}(r_1 = 0) < 1 + \theta$, namely:

$$F\big(\hat{c_2}(r_1 = 0)\big) = \frac{1+\theta}{\bar{c}} \tag{2}$$

The decision of P_1 depends on their beliefs about P_2's report. P_1 lies if $E(U_1(r_1 = 1)) > E(U_1(r_1 = 0)$. Let b_0 be P_1's belief that P_2 reports 1 after $r_1 = 0$, and b_1 be P_1's belief that P_2 reports 1 after $r_1 = 1$. Then, taking into account the utility presented in (1), P_1 lies if $1 - c_i + \theta(1 - b_1) > b_0$. This implies that the lying threshold that divides those who lie from those who do not in AVOID is:

$$\hat{c_1} = 1 - b_0 + \theta(1 - b_1) \tag{3}$$

In equilibrium, the beliefs about the response of P_2 are $b_1 = 0.2$, given that this is the probability of drawing 1, and $b_0 = 0.2 + 0.8\frac{1+\theta}{\bar{c}}$. Thus, replacing b_1 and b_2 in equation $\hat{c_1}$, I get that the lying threshold at equilibrium for P_1 in AVOID is:

$$\hat{c_1} = 0.8\left(1 + \theta - \frac{1+\theta}{\bar{c}}\right) \tag{4}$$

2.2.2 Treatment 2: NO AVOID

In a second treatment, NO AVOID, I remove the P_1's capacity of relying on P_2 possibility to lie by imposing $x_2 = r_2$. To do so, in NO AVOID, participants with the role of P_2 do not have the possibility of reporting their random draw, but the computer will record the random draw and report it truthfully. This procedure is common knowledge. The payoff structure is the same as in AVOID, and even if a computer makes the report, a human participant bears the consequences in terms of payoffs. Thus, with this procedure, I ensure that P_1 has an objective probability of r_2. This feature implies that $b_1 = b_0 = 0.2$. Then, using (3) and substituting the new values of b_1 and b_2, the threshold of the lying cost at equilibrium for P_1 in NO AVOID is:

$$\hat{c_1} = 0.8\,(1+\theta) \tag{5}$$

Comparing the lying cost thresholds presented in (4) and (5), it follows that more P_1 will lie when they can not rely on P_2's incentives to lie. This result holds because the utility by prosociality and the maximum lying cost are non-negative.

Hypothesis 1 (No cost avoidance)**.** *In* NO AVOID*, the proportion of* P_1 *who lie will be higher compared with* AVOID*.*

2.2.3 Treatment 3: NO EXTERNALITY

In a third treatment, I investigate the role of the positive externality on P_1's decision. I use the same structure as in NO AVOID but change the payoff scheme to eliminate the benefit on others, so that in this treatment lies are no longer Pareto White Lies but pure selfish lies. I keep P_1's monetary payoffs identical as in NO AVOID but make P_2's monetary payoffs only dependent on x_2. In particular, in NO EXTERNALITY, P_2 gets 1 only if $x_2 = 1$ and 0 otherwise. The variation in this treatment implies that in the utility function $\theta = 0$. Therefore, the threshold of the lying cost at equilibrium for P_1 in NO EXTERNALITY is:

$$\hat{c_1} = 0.8 \tag{6}$$

From the comparison between (5) and (6) it follows that lying is more pronounced in NO AVOID than in NO EXTERNALITY.

Hypothesis 2 (Positive externality)**.** *In* NO EXTERNALITY*, the proportion of* P_1 *who lie will be less compared with* NO AVOID*.*

Given that in NO EXTERNALITY lying decreases compared with NO AVOID, one question remaining is whether NO EXTERNALITY accounts for the same effect than AVOID. However, this effect depends on $\bar{c}$. When $\bar{c}$ is lower than 1, lying will be higher in NO EXTERNALITY than in AVOID. This means that the comparison in lying rates between AVOID and NO EXTERNALITY will depend on the proportion of people with high lying cost in the population.

Until this point, I have presented the main treatments that allow me to assess the impact of lying aversion and prosocial lying on the preferences to lie. To sum up, in NO AVOID the probability of P_2 reporting 1 is fixed at 0.2 (in contrast to AVOID where it was a subjective probability). So, P_1 has more room to avoid the lying cost in AVOID than in NO AVOID, but the prosocial motive is still present in both conditions. Therefore, NO EXTERNALITY lets me assess the role of prosocial lying. Table 1 illustrates how the experimental design isolates each potential explanation allowing me to assess each motive.

Table 1. Comparison of lying aversion and prosociality across sequential treatments.

	AVOID	NO AVOID	NO EXTERNALITY
Avoid the Lying Costs $P(r_2 = 1 \vert r_1 = 0)$	$0.2 + b_0$	0.2	0.2
Prosociality	✓	✓	×

Note: the row *Avoid the Lying Cost* refers to how likely is to effectively avoid the lying cost while getting the high payoff. For AVOID it uses b_0 to represent the subjective probability P_1 attributes to P_2 reporting 1.

2.2.4 Treatment 3: SIMULTANEOUS

The last treatment, SIMULTANEOUS, uses the same payoffs structure as in AVOID but participants report simultaneously instead of sequentially. Playing sequentially allows P_1 to transmit their action r_1 to P_2 and gives some strategic advantage to P_1. In contrast, in SIMULTANEOUS participants need to act without any information about the partner's actual decision. In the utility function presented in (1), I assume that lies that benefit others generate some utility represented by θ. However, θ only counts if the benefited individual does not report 1. This assumption implies that individuals use consequentialist norms when lying for others, where the action itself does not matter but only the consequence on others payoffs.

In SIMULTANEOUS, players are symmetric and no information is learned before deciding. Hence, I do not use b_0 and b_1, but define b_{ij} as the belief of P_i that P_j reports 1. As in AVOID, P_i lies if $E(U_i(r_i = 1)) > E(U_i(r_i = 0))$. That is, P_i lies if $1 + \theta(1 - b_{ij}) - c_i > b_{ij}$, which leads to the lying threshold $\hat{c_i} = (1 - b_{ij})(1 + \theta)$. In equilibrium, $b_{ij} = 0.2 + 0.8\frac{\hat{c_i}}{\bar{c}}$. Thus, I plug b_{ij} in the threshold equation to get $\hat{c_i} = (0.8 - 0.8\frac{\hat{c_i}}{\bar{c}})(1 + \theta)$. It follows that the lying threshold in Simultaneous is:

$$\hat{c_1} = 0.8\left(\frac{0.8\bar{c}(1+\theta)}{1+0.8(1+\theta)}\right) \tag{7}$$

The resulting lying threshold in Simultaneous presented in (7) needs to be compared with (4). However, this comparison is not as trivial as in the other treatments. Individuals have two competing motives when deciding whether to lie in Simultaneous and Avoid. On the one hand, they hope to be able to rely on their partner's incentives and avoid the psychological cost of lying. On the other hand, they have the prosocial motive when lying for others and then can use it to decrease their cost of lying. The first motive, lying aversion, implies that P_1's motive to lie out of own-payoff consideration is stronger in Simultaneous than in Avoid because of sequentiality. However, it is more difficult for P_i to use the prosocial motive in Simultaneous than in Avoid because consequential prosocial lying implies that reporting 1 only increases individuals' payoffs if their partner reports 0. In other words, in Simultaneous an individual P_i may be willing to lie and use the prosociality to decrease the cost of lying, but P_j is likely doing the same, and none of them gets θ which can be anticipated for both players and leads to no one lying.

By observing the lying thresholds in (4) and (7), one can see that determining which motive dominates the other depends on the combination of θ and $\bar{c}$. To understand this relation, I calculate numerically the values of θ and $\bar{c}$ that imply the same lying rates in Avoid and Simultaneous. Figure 1 shows that when θ is high the prosocial motive is stronger than the cost avoidance motive.[5] In this case, lying is higher in Avoid than in Simultaneous. Conversely, if θ is low enough, it is more likely that the cost avoidance motive plays a central role, and thus lying would be higher in Simultaneous than in Avoid (shaded area in Figure 1). As Figure 1 shows, lying can be higher or lower in Avoid compared with Simultaneous; therefore, I will test the hypothesis that individuals will lie more in Simultaneous than in Avoid under the conjecture that the motive of lying aversion dominates the motive of prosocial lying.[6]

Hypothesis 3 (No cost avoidance simultaneous)**.** *In* Simultaneous*, the proportion of* P_1 *who lie will be more compared with* Avoid.

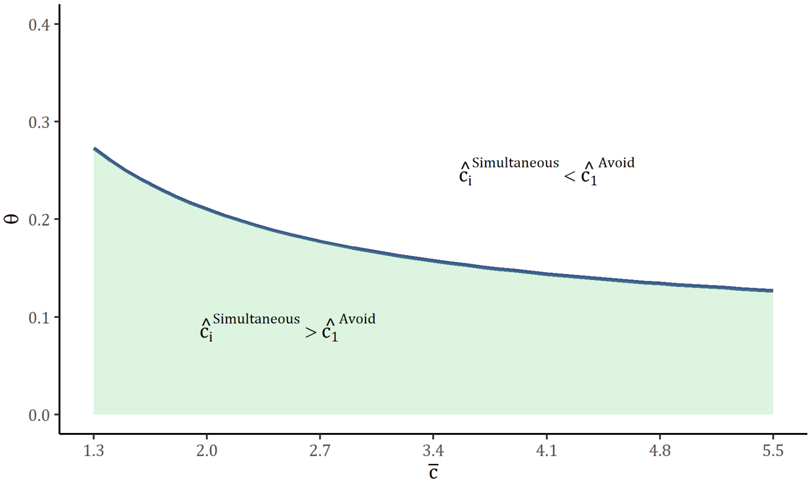
Table 2 summarizes the decisions each player has to make across the four conditions, the payoff function, and the hypotheses based on the model. I will use this experimental

---

[5] Note that a value of 0.3 means that the utility generated by the positive externality is equal to a 30% of the utility generated by the monetary payoff.

[6] All the hypotheses were preregistered in the AEA RCT Registry.

Figure 1. Comparison of Lying Thresholds in Avoid and Simultaneous



design in two different experimental studies where the same variations apply, and I will test the same hypotheses. It is also important to note that direct treatment comparisons are only possible in the following pairs: Avoid-No Avoid, Avoid-Simultaneous, No Avoid-No Externality.

Table 2. Summary of the actions, payoffs and hypotheses in each treatment

	P_1	P_2	Payoffs	H_0
Avoid	reports r_1	learns r_1 and then reports r_2	$v_i = \begin{cases} v_l = 0 & \text{if } r_i = r_j = 0 \\ v_h = 1 & otherwise \end{cases}$	-
No Avoid	as in Avoid	learns r_1 but the report is made by the computer.	as in Avoid	P_1 lies more than in Avoid
No Externality	as in No Avoid	as in No Avoid	$v_1 = \begin{cases} v_l = 0 & \text{if } r_i = r_j = 0 \\ v_h = 1 & otherwise \end{cases}$ $v_2 = \begin{cases} v_l = 0 & \text{if } r_2 = 0 \\ v_h = 1 & otherwise \end{cases}$	P_1 lies less than in No Avoid
Simultaneous	Players make simultaneous decisions		as in Avoid	P_1 lies less than in Avoid

3 Experimental Studies

3.1 Study 1: Two person cheating game – Observed

3.1.1 Overview and design Study 1

In Study 1, for the random draw, participants click on a card that reveals a color. There are two possible colors. Reporting $Orange$ pays $£4$, reporting $Black$ pays $£0.5$. The probability of drawing $Orange$ is 0.2 and $Black$ 0.8. In this study, given that the random draw x_i is observed by the experimenter, it is possible to identify whether P_1 lied or not.

In AVOID, P_1 clicks on a box in the computer screen and a color (Orange or Black) is revealed. Participants know that the probability of drawing $Orange$ is 0.2 and $Black$ 0.8. After P_1 observes the drawn color, they are asked to report the color to P_2. Once P_2 learns what P_1 has reported, they are asked to click on a box in the computer screen that reveals a color (Orange or Black). Then, P_2 reports their observed color. In NO AVOID, P_1's decisions are the same, but P_2 only clicks on the box they see on the computer screen, and the computer reports truthfully the random draw. In NO EXTERNALITY, decisions are identical to NO AVOID and the variation is on P_2's payoffs which only depends on their own random draw. Finally, in SIMULTANEOUS, participants make the same decision as P_1 in AVOID but both members of the dyad decide without knowing the other's report r_j.

While P_2 is reporting, I elicit P_1's beliefs about P_2's report. I use a mechanism proposed by Karni (2009) and implemented experimentally first by Mobius et al. (2011) which allows eliciting probabilities in an incentive-compatible way. Specifically, I use a similar implementation as the one proposed by Coffman (2011). Participants are asked to guess the color reported by P_2 and then ask how likely they think their guess is correct. This procedure allows me to elicit the probability of the P_2 reporting $Orange$. Participants are told that they do not need to read the instructions about the mechanism or understand it if they do not want to. I use this option to reduce the risk of people leaving because of the complexity of the mechanism. They can, however, click on a button to see the detailed explanation.[7]

Specifically, the elicitation mechanism is based on robots that can guess on behalf of the participants. There are 100 robots, each with integer probability between 1 and 100 of correctly guessing P_2's report. A robot from this interval is drawn randomly, and it can guess on the participant's behalf with an accuracy level determined by its number. Robot 1 is accurate 1% of the time; robot 2 is accurate 2% of the time, all the way up to the robot that is accurate 100% of the time. The reported likelihood of their their guess being correct is used as an "accuracy threshold." That is, if the robot has an accuracy greater than or equal to the threshold, the robot guesses r_2 for P_1. If the robot has an accuracy less than the threshold, P_1's guess is submitted. If the guess is correct, whether it is the participant's or the robot's, it gives a payoff of $\pounds 0.5$.

---

[7] From the total of participants in P_1 role, 33.46% clicked once in the info button, and 0.38% clicked twice.

3.1.2 Procedures Study 1

I pre-registered the experiment in AEA RCT Registry under the number AEARCTR-0006881. I targeted 120 observations per treatment ex-ante. I calculated the power of the target sample size using computer simulations. I used a minimum detectable effect size of 0.15 from people detected as liars. The power reached with a sample size of 120 observations by treatment is 0.8 when simulating 1500 Fisher tests. The experiment was conducted online on Prolific (Palan and Schitter, 2018) in December 2020. The experiment was programmed in oTree (Chen et al., 2016). A total of 878 people participated in five sessions.[8] I did not run the whole experiment in one session to avoid overloading the server and minimize the probability of technical issues. Table 3 presents the number of valid observations for people on the role of P_1 in each session. The computer program assigned a treatment to each participant. Participants participated only in one treatment, and the game was played only one time. Among the participants, 51.17% identified themselves as male, 47.27% as female, 0.89% as other, and 0.66% preferred not to report it. The average age of the participants was 26.06, and 48.53% were students.

Table 3. Participants with the role of P_1 in Study 1

	Session 1	Session 2	Session 3	Session 4	Session 5	**Total**
AVOID	22	23	22	23	35	125
NO AVOID	22	21	21	22	35	121
NO EXTERNALITY	23	24	22	21	37	127
SIMULTANEOUS	20	22	22	20	48	132

Participants read the instructions first and responded to some comprehension questions. After they answered the comprehension questions correctly, they waited until a second player was also ready, then they were matched together and proceeded to the observed cheating game. Roles were assigned randomly. After participants finished the observed cheating game and the elicitation task, they responded to a survey with demographic questions and a feedback question. Participants spent about 6 minutes on average to complete the experiment. Additional to the earnings on the cheating game and the guessing task, participants earned a completion fee of $\pounds 2.5$. Following Prolific rules, participants who left the experiment did not get the completion fee. Participants received their payoffs through the Prolific platform the same day they participated in their session.

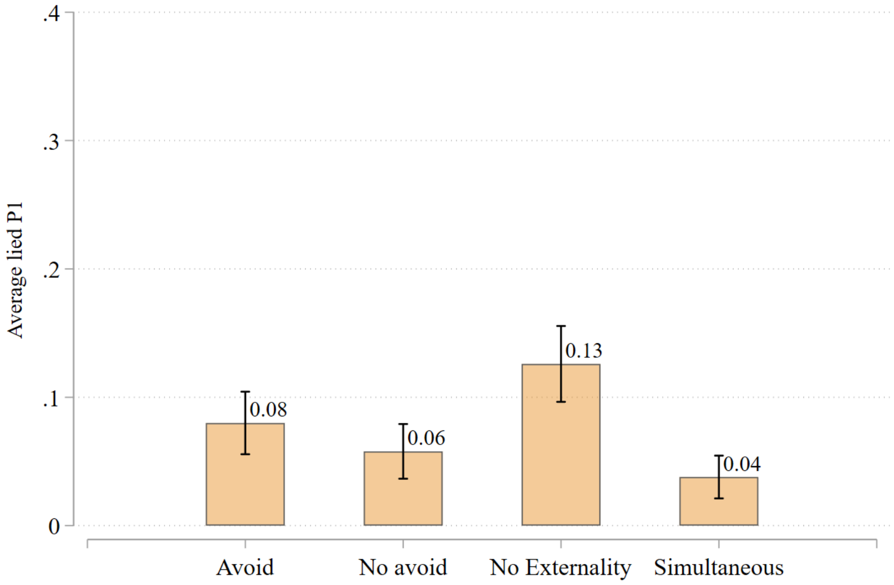
---

[8] A total of 899 people showed up, but some left in the middle of the session.

3.1.3 Result Study 1

The main outcome variable of interest to test the hypotheses presented in Section 2 is whether P_1 lied or not. Given that in this experiment I have the information about the random draw and the report, I create a new variable called *Lied* that takes the value of 1 when $x_i \neq r_i$, and 0 when $x_i = r_i$. Figure 2 presents the lying rates per treatment. The lying rates are on average 7.52%, which is very low compared, for instance, with Gneezy et al. (2018) where lying rates were about 30%. The pairwise comparison of the comparable treatments using a Fisher exact test results in no significant differences at a 0.05 level across treatments.[9]

Figure 2. **P_1's lying rates across treatments in Study 1**



To confirm the result implied by Figure 2, I use regressions that allow me to include some controls. Table 4 presents, in columns 1 and 2, Linear Probability regressions with *Lied* as dependent variable. Lied 1 uses data from treatments AVOID, NO AVOID, and SIMULTANEOUS, with AVOID as the reference treatment. Lied 2 uses data from treatments AVOID, NO AVOID, and NO EXTERNALITY, with NO AVOID as the reference treatment. I use two regressions because AVOID is not directly comparable with NO EXTERNALITY which make it impossible to include NO EXTERNALITY in Lied 1. The same intuition applies to Lied 2 because between NO AVOID and SIMULTANEOUS two things change. In both regressions, I use only the data of P_1 that drew $Black$. The independent variables are the treatment dummies, the beliefs about r_2[10], the time participants took to send the report, some demographic variables, and their id in the session to control for potential selection between the first participants that entered the session and the last ones.

---

[9] To ensure that only one component changes, I only compare treatments in the following pairs: AVOID-NO AVOID, AVOID-SIMULTANEOUS, and NO AVOID-NO EXTERNALITY.

Table 4. Regressions testing the differences across treatments in Study 1

	Lied 1	Lied 2	Beliefs 1	Beliefs 2
Avoid	*Reference*	-0.005	*Reference*	7.875***
		(0.035)		(2.811)
No Avoid	0.012	*Reference*	-7.607***	*Reference*
	(0.036)		(2.755)	
No Externality		0.052		3.183
		(0.045)		(2.591)
Simultaneous	-0.024		-1.191	
	(0.035)		(2.965)	
Belief about $Pr(r_2 = 1)$	0.004***	0.002**		
	(0.001)	(0.001)		
Time Spent Reporting	0.008***	0.012***		
	(0.003)	(0.004)		
Lied=1			24.105***	9.199*
			(5.288)	(4.704)
Constant	-0.136**	-0.095	33.189***	26.030***
	(0.058)	(0.081)	(4.294)	(3.878)
Controls	Yes	Yes	Yes	Yes
Observations	302	292	302	292
R^2	0.133	0.098	0.132	0.081

Note: Regressions Lied 1 and Lied 2 are Linear Probability models. Regressions Beliefs 1 and Beliefs 2 use OLS. All regressions use the data of participants with the role of P_1 and who drew $Black$. Lied 1 and Belief 1 use data from treatments AVOID, NO AVOID, and SIMULTANEOUS. Lied 2 and Beliefs 2 use data from treatments AVOID, NO AVOID, and NO EXTERNALITY. Controls include gender, age, student status, education, number of experiments they participated in before, and their id in a session. Robust standard errors are presented in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

The coefficients for the treatments dummies in Lied 1 and Lied 2 confirm no significant differences in lying rates across treatments. Additionally, I find that participants that spent more time reporting were more likely to lie. Interestingly, the coefficient of *Time Spent Reporting* shows that the probability reported by P_1 is positively correlated with the probability that $r_1 = 1$.

---

[10] I take the confidence on the guess (accuracy threshold) of each P_1. In case they guessed $x_2 = Black$, the value of the variable is $1 -$ *accuracy threshold*.

Result 1 (No treatment differences on lying)**.** *There are no treatment differences between treatments leading me to reject Hypotheses 1, 2, and 3. This result may, however, at least partially, be driven by low lying rates.*

The mean beliefs by treatment are not significantly different in all pairwise comparisons using a Kolmogorov-Smirnov test. In columns Beliefs 1 and Beliefs 2 of Table 4 I use a Ordinary Least Squares to assess whether the reported probability of reporting $Orange$ varies across treatments. Belief 1 uses data from treatments AVOID, NO AVOID, and SIMULTANEOUS with AVOID as the reference treatment. Beliefs 2 uses data from treatments AVOID, NO AVOID, and NO EXTERNALITY with NO AVOID as the reference treatment. Beliefs 1 show that the reported probability in NO AVOID is lower than in AVOID while the difference between AVOID and SIMULTANEOUS is non-significant. Column Beliefs 2 shows that the difference between NO AVOID and NO EXTERNALITY is non-significant neither.

3.2 Study 2: Two person cheating game – Mind

3.2.1 Overview and design Study 2

One potential reason why Study 1 presented minimal lying rates, making detecting treatment differences difficult, is that people might be concerned about the random draw's observability. Gneezy et al. (2018) and Fries et al. (2021) show that, in laboratory experiments, the observability of the random draw decreases lying. In Study 1, maybe because it was an online experiment, and given that Prolific emphasizes on the importance that their participants respond everything honestly, this effect was exacerbated.[11] To assess whether behavior changes when lying is not observed, I designed a second study where the random draw is private and not observed, not even by the experimenter.

In this second study, I use a mind-cheating game[12] in which participants choose one color out of five in their minds (see colors in Figure 3).[13] Then, they draw a color from a deck of cards presented on their computer's screen. The deck of cards contains two cards for each one of the colors. Participants then report whether the color they drew from the deck is the same as their mentally chosen color. If participants want to report that the colors match, they report *Yes*; otherwise, they report *No*. Thus, in this second game, *Yes* represents $x_i = 1$ and *No* represents $x_i = 0$. For the payoffs, the rewards I use are $v_h = £2.5$ and $v_l = £0.3$.[14] In this study, the state of nature is in participants' minds, so I can only compare distributions of groups based on the known theoretical distribution. However, I cannot identify whether an individual lies or not. Furthermore, the probability of having matching colors is the same as drawing $Orange$ in Study 1.

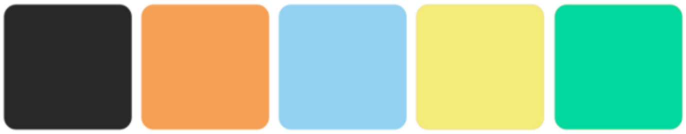
---

[11] For instance, in the first study every participant has to complete before participating in further studies in Prolific, they include the following statement: "*...we want to build a world where people and organisations can make important decisions based on trustworthy data and solid evidence. We can't build that world without your contribution: The data you provide, combined with your honesty, your integrity and your effort, is a precious piece of the research puzzle. And together, those pieces help advance human knowledge.*"

[12] Mind games were previously implemented using die rolls (Jiang, 2013; Shalvi and De Dreu, 2014; Potters and Stoop, 2016; Kajackaite and Gneezy, 2017; Dimant et al., 2020) or coin tosses (Shalvi et al., 2012; Garbarino et al., 2019). One potential flaw of traditional mind games which use die rolls is that they could be biased if people prefer certain numbers, and then the experimenter loose control about the theoretical distribution of the random draws.

[13] The colors were chosen such that people with colorblindness can see five different colors.

Figure 3. Colors used in mind game



I use the same treatments presented in Table 2. Specifically, in AVOID, P_1 reports to P_2 whether the colors match or not.[15] Once P_2 learns r_1, they follow the same sequence of decisions: think a color, draw a color from a deck of cards, and report whether the colors match. In NO AVOID, P_1's decisions are the same, but P_2 do not select their card in their mind, but they selected it from a list presented on their screens. Then, they draw a color from a deck of cards. Finally, using the selected color and the drawn color, the computer reports whether the colors match or not. In NO EXTERNALITY, decisions are identical to NO AVOID, and the variation is that P_2's payoffs only depend on whether their selected color and their drawn color match regardless of P_1's payoffs. Finally, in SIMULTANEOUS, both participants think of a color, draw a color, and report at the same time whether the colors match. I also elicit $P_2 1$'s beliefs in all the treatments using the same mechanism as in Study 1 and paying $£0.3$ if they guess correctly. Finally, in this second study, I included information on participants' religion in the demographic variables.

3.2.2 Procedures Study 2

Procedures in Study 2 are the same as in Study 1. I pre-registered the experiment in AEA RCT Registry under the number AEARCTR-0007214. I created a new pre-registration entry because this study is not a modification of Study 1, but a new study to check whether having a non-observable random draw changes participants' response to incentives. Therefore, I decided to pre-register this study in a separate register before running its sessions. A total of 992 people participated in five sessions.[16] Table 5 presents the number of observations for people on the role of P_1 in each session. Among the participants, 54.71% identified themselves as male, 44.60% as female, 0.30% as other, and 0.40% preferred not to report it. The average age of participants was 26.25, and 47.28% were students. Participants spent about 7 minutes on average to complete the experiment. In addition to the mind game earnings and the guessing task, participants earned a completion fee of $\pounds 1.15$.

---

[14] I used lower payoffs because after running the first study, I realized that I was paying a lot compared with other Prolific studies, and I also wanted to rule out participants being positively reciprocal to the experimenter.

[15] One advantage of the procedure I use is that independent of the color chosen by participants, the probability of matching is always 0.2, which is the same as drawing $Orange$ in Study 1.

Table 5. Participants with the role of P_1 in Study 2

	Session 1	Session 2	Session 3	Session 4	**Total**
Avoid	29	38	38	37	142
No Avoid	29	36	39	37	141
No Externality	28	37	39	36	140
Simultaneous	30	38	40	38	146

3.2.3 Results Study 2

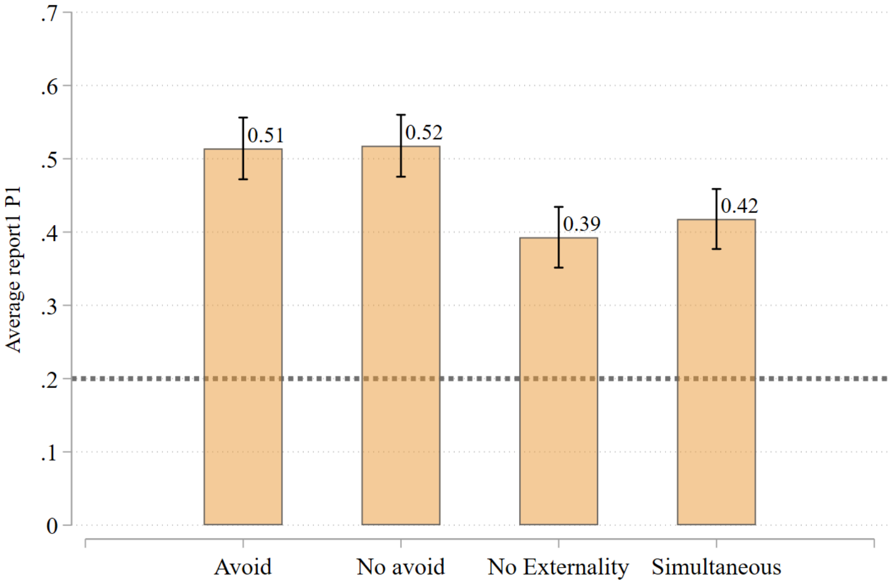
Given that I used a mind game in this study, the 'state of the world' is private information, and I can only compare the reports at an aggregate level. Theoretically, the random draw follows a binomial distribution with a probability of the high-paying state of 0.2. Figure 4 shows the proportion of those participants with the role P_1 that reported *Yes* and those who reported *No*. Using the Binomial test, I confirmed that the actual reports are statistically different from the reports under full honesty in all treatments. I calculate the expected lying rates of the reports in Figure 4 by taking the average of the reports, then subtracting 0.2 (the expected proportion of people actually matching colors), finally I divide the result over 0.8. The expected lying rates are 38.75% in AVOID, 40% in NO AVOID, 23.75% in NO EXTERNALITY, and 27.50% in SIMULTANEOUS. The pairwise comparisons using one-sided Fisher Exact test show that the difference between AVOID and NO AVOID is not significant ($p = 0.523$), the difference between NO AVOID and NO EXTERNALITY is significant ($p = 0.024$), and the difference between AVOID and SIMULTANEOUS is weakly significant ($p = 0.064$).

I use a Linear Probability Model estimation to assess the treatment effects once I control for potential confounding variables. In columns 1 and 2 of Table 6, I present two regressions with r_1 as the dependent variable. The regressors are the treatment dummies, the beliefs about r_2, the time participants took to report, some demographic variables, and

---

[16] A total of 1009 people showed up, but some left in the middle of the session

Figure 4. P_1**'s *Yes* reports across treatments in Study 2**



Note: The dashed horizontal lines display the underlying theoretical proportion of *Yes* under truth-telling.

their id in the session. The regression *Report* P_1 *1* confirms the result derived from Figure 4 that there is no difference in lying between AVOID and NO AVOID leading to Result 2.

Result 2 (Related to Hypothesis 1)**.** P_1 *lying behavior is not different in* AVOID *than in* NO AVOID. *I reject Hypothesis 1.*

Also in regression *Report* P_1 *1* the coefficient Simultaneous shows that, once I control for potentially confounding variables, in AVOID P_1 lied significantly more than in SIMULTANEOUS. This finding is in the opposite direction of hypothesis 3 which states that lying will be more pronounced in SIMULTANEOUS than in AVOID.

Result 3 (Related to Hypothesis 3)**.** P_1 *lies more in* AVOID *than in* SIMULTANEOUS. *I reject Hypothesis 3.*

Regression *Report* P_1 *2* in Table 6 uses as a reference group NO AVOID because this is directly comparable with AVOID and NO EXTERNALITY. This regression confirms the null effect in lying between NO AVOID and AVOID. More importantly, this regression reveals that P_1 lying is lower in NO EXTERNALITY than NO AVOID which is consistent with hypothesis 2.

Result 4 (Related to Hypothesis 2)**.** P_1 *lies less in* NO EXTERNALITY *than in* NO AVOID. *I do not reject Hypothesis 2.*

Table 6. Regressions testing the differences across treatments in Study 2

	Report P_1 1	Report P_1 2	Report P_2 AVOID
Avoid	*Reference*	-0.015	
		(0.060)	
No Avoid	0.033	*Reference*	
	(0.060)		
No Externality		-0.123^{**}	
		(0.060)	
Simultaneous	-0.142^{**}		
	(0.061)		
Belief about $Pr(r_2 = 1)$	0.003^{***}	0.002	
	(0.001)	(0.001)	
Time Spent Reporting	-0.004	-0.001	-0.008^{***}
	(0.003)	(0.003)	(0.003)
P_1's report=1			-0.228^{***}
			(0.082)
Constant	0.351^{***}	0.340^{***}	0.734^{***}
	(0.116)	(0.122)	(0.176)
Controls	Yes	Yes	Yes
Observations	428	421	142
R^2	0.051	0.031	0.121

Note: Regressions Report P_1 1, Report P_1 2, Report P_2 are Linear Probability models. Report P_1 1 uses data from treatments AVOID, NO AVOID, and SIMULTANEOUS. Report P_1 2 uses data from treatments AVOID, NO AVOID, and NO EXTERNALITY. Report P_2 AVOID uses data from P_2 in AVOID. Controls include gender, age, student status, education, religion, number of experiments they participated before, and their id in a session. Robust standard errors are presented in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

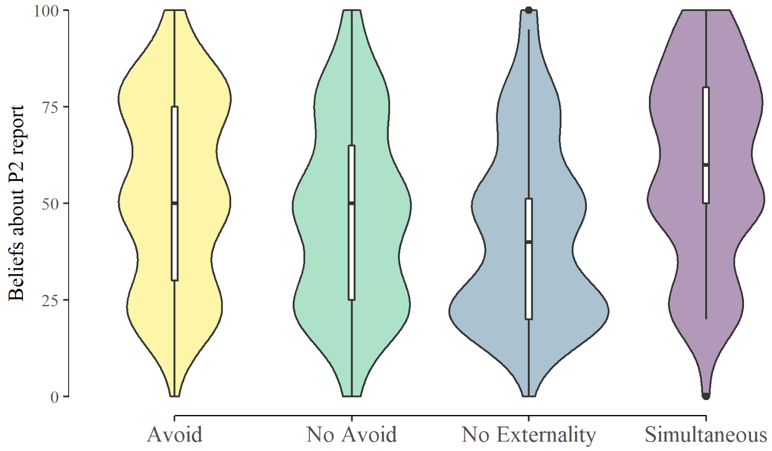
Besides the results concerning P_1's reports, Table 6 also presents important evidence regarding P_2 reports. In the column *Report* P_2, it presents the relation between r_1 and r_2 controlling for the time participants take when reporting and demographics variables. In this regression, I only used data from AVOID because it is the only treatment where P_2 can lie after learning r_1. The coefficient P_1*'s report=1* shows that P_2 was significantly more likely to report *Yes* when $r_1 =$ *No*. Using Hugh-Jones (2019) Bayesian method, I estimate the lying rates of P_2 conditional on r_1. The lying rate when P_2 observed $r_1 =$ *Yes* is 9.39% while the lying rate when observing $r_1 =$ *No* is 36.63%.

Result 5. *P_2 lie more when they observe that P_1 reported No than when they observe that P_1 reported Yes.*

Figure 5 presents the distribution of the implied probabilities of P_2 reporting *Yes* in each treatment. I use a Kolmogorov-Smirnov test to test whether these distributions are equal.

The pairwise comparisons using this test shows that in the pairs AVOID-SIMULTANEOUS and NO AVOID-NO EXTERNALITY there is no statistically significant difference, but in the pair AVOID-NO AVOID there is a weakly significant difference pointing to smaller numbers in NO AVOID ($p = 0.084$).

Figure 5. P_1**'s subjective probability that** P_2 **reports** ***Yes***



Note: The graphs uses a kernel density plot on each side. Inside there is a box-plot.

To complement the insights from Figure 5, I use OLS regressions to study the differences in beliefs across treatments. Table 7 presents two regressions (one for each reference treatment) with the belief about P_2 reporting *Yes* as the dependent variable. I also include interaction terms of the report of P_1 and each treatment. Regression *Beliefs 1* of Table 7 shows that those participants in simultaneous who reported *No* believed that it was more likely that their partner would report *Yes*. Table 7 also shows that beliefs were not self-serving in the mind game.

Result 6 (Beliefs in mind game)**.** *P_1's subjective probability of $r_2 = 1$ is not positively correlated with r_1 in the mind game. Participants in* SIMULTANEOUS *believe that it is more likely that their partner reports Yes than in other treatments.*

Finally, in Study 2, I included a question in the final questionnaire where I asked them: "*Imagine you were to play the same game again and had a choice, would you rather be Participant A or Participant B* [17]" Their responses, divided by the role they had, are presented in Figure 6. Figure 7(a) shows that P_1 did not interpret, in general, being the first mover as an advantage in AVOID or SIMULTANEOUS. Figure 7(b) shows that this is also true for P_2 with even more people being completely indifferent between the two roles. Figure 7(b) also shows that P_2 did not like the second mover position when they could not report and

---

[17] This was the exact wording I used in both experiments to refer to P_1 and P_2.

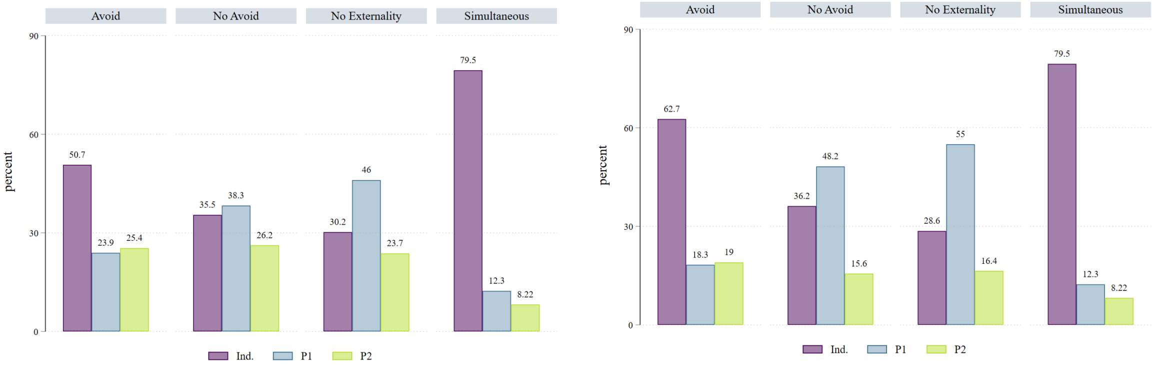
Table 7. Regressions testing the differences in Beliefs across treatments in Study 2

	Beliefs 1	Beliefs 2
Avoid	*Reference*	3.711
		(4.042)
No Avoid	-3.719	*Reference*
	(4.025)	
Simultaneous	7.895^{**}	
	(3.691)	
No Externality		-3.379
		(3.721)
P_1's report=1	6.397	2.297
	(4.090)	(4.102)
Avoid × P_1's report=1		4.129
		(5.791)
No Avoid × P_1's report=1	-4.901	
	(5.785)	
Simultaneous × P_1's report=1	-2.963	
	(6.192)	
No Externality × P_1's report=1		-0.439
		(5.834)
Constant	49.835^{***}	55.810^{***}
	(5.220)	(5.499)
Controls	Yes	Yes
Observations	428	421
R^2	0.055	0.051

Note: Regressions Report P_1 1, Report P_1 2, Report P_2 are Linear Probability models. Controls include gender, age, student status, education, religion, number of experiments they participated before, and their id in a session. Robust standard errors are presented in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

would prefer being P_1.

Figure 6. Which role would participants choose if they play again?



(a) P_1 (b) P_2

4 Discussion

The present study was designed to determine whether prosociality or lying aversion have more weight in individuals' preferences for lying. The first experiment, Study 1, only could provide limited evidence in this respect because participants were reluctant to lie in the observed cheating game. Previous evidence of laboratory experiments on cheating also shows that in observed cheating games, people lie less (Gneezy et al., 2018; Abeler et al., 2019; Fries et al., 2021). However, they were able to detect treatment effects. I used this fact to design Study 1 and get information at the individual level. In contrast with laboratory experiments, the online setting provided lower lying rates making difficult to detect treatment variations.

To make it easier to detect treatment differences, I used a mind game where it was not possible to know whether a specific participant lied or not, but the random draw was only on participants' minds. The main advantage of the observed cheating game was that I could identify liars and get more information about liars and no liars. In contrast, participants are more willing to lie in the mind cheating game, but I can detect lying only at the aggregate level. In Study 2, where no observability of the random draw was delivered to participants, it was possible to assess the hypotheses presented in Section 2 because participants were more sensitive to incentives. So, in this section, I will mainly discuss the results from Study 2.[18]

---

[18] The main lesson when comparing the behavior from Study 1 and Study 2 is that it is crucial to reduce observability the most when using platforms such as Prolific (Palan and Schitter, 2018) to study lying behavior. A potential reason is that they stress the importance of being honest when participating in studies. Then, even

I hypothesized that most of participants with the role of P_1 in the treatment AVOID would try to save their lying costs and pass the burden to P_2. Result 2 shows that this is not the case, and participants had lying rates similar to NO AVOID. This finding was unexpected and suggests that either P_1 expected that P_2 will not lie or that the utility derived from the positive externality was stronger than the direct cost of lying. The beliefs of P_1 in AVOID, presented in Figure 5 and Table 7, rule out the possibility that, in AVOID, most of P_1 believed that P_2 is particularly honest. Additionally, Result 5 shows that people are willing to avoid the direct costs of lying when their actions will not have implications on others' payoffs or their payoffs. Result 4 shows indeed that when there is no prosocial motive, P_1 lied considerably less.

Hence, the most important insight from the paper is that even if people are lying averse, when they benefit others, prosociality is a stronger motivator than lying aversion. This result adds to the previous finding by Wiltermuth (2011), Gino et al. (2013), and Levine and Schweitzer (2015) by showing not only that individuals are more willing to lie when they create a positive externality but also that this motive is even stronger than the aversion generated by the psychological lying cost. Now the question is how SIMULTANEOUS enters to this picture. Arguably, SIMULTANEOUS provides two contributions to the main result of the paper. First, it confirms the validity of the assumption regarding prosociality. Specifically, I used a particular assumption about the type of prosocial preferences where participants benefit from the consequences of their actions and not by the actions themselves. These type of prosocial preferences are inspired by Consequentialism and juxtaposed to Deontological ethics that evaluate the means instead of the ends. Even though SIMULTANEOUS was not intended to test this assumption, Result 3 provides evidence for participants holding consequentialist prosocial preferences.

Second, and more importantly, SIMULTANEOUS confirms that the prosocial motive is strong enough to dominate the lying aversion motive. As presented in Figure 1 of section 2, when the utility generated by the positive externality (θ) is high enough, lying is higher in AVOID than in SIMULTANEOUS. This finding was not expected because I predicted that people would be more self-interested and avoid the psychological cost of lying more than trying to help others. However, once we establish that the prosocial motive dominates the lying aversion motive, this result is consistent with the model.

Finally, regarding the elicited beliefs about r_2, in Study 2 participants in SIMULTANEOUS were more likely to believe that their partner will report *Yes* (see Result 7). This finding is consistent with the model. It implies that it will be easier to avoid the lying cost but also that it is more likely to lose the utility of the positive externality given that they are

---

if the only identification is their Prolific ID, participants may care about how they are perceived.

consequentialists. By contrast, in Study 1, the main result concerning beliefs was that they were self-serving. It could be argued that this result was due to participants trying to justify themselves when lying.

5 Conclusion

Is prosocial lying a stronger motivation than lying aversion? In the setting of this paper, there is evidence that this is the case. Studies of Gneezy et al. (2018), Abeler et al. (2019), and Khalmetski and Sliwka (2019) have indicated that one explanation for people not lying to maximize their monetary payoff is that they have psychological lying cost. It is not clear, however, that this effect holds in group settings where prosocial lying enters into the picture. For instance, it is not the same lying to get an individual incentive a CEO has promised when achieving a goal than lying to reach a threshold that gives a bonus for a team of workers. A similar dilemma occurs when people use intermediaries in some situations, such as tax declarations, selling a car, or selling a house. In these contexts, the intermediary has a higher payoff if they lie. Therefore, when individuals lie to benefit others as well as themselves, there are two competing motivators: lying aversion and prosociality.

I used two online experiments to study these situations. I found meager lying rates in the first experiment, making it difficult to identify any treatment differences. Arguably, the main reason for the low lying rates was that I used an observed game, which makes it possible to know whether a participant lied. In a second experiment, I solved the observability problem by using a mind game. Study 2 presented evidence that individuals lie more when they can benefit others. In addition, it showed that this motivation is strong enough to prevail even when people could save their direct lying costs.

One additional finding was that prosocial lying is consistent with consequentialism rather than deontological views. In particular, prosociality being a stronger motivator than lying aversion is only possible in the theoretical model if people care about the consequences of their acts rather than their intentions. However, the scope of this study was limited in terms of assessing whether consequentialist prosocial behavior is the only way to explain the results, and the experiment was not designed to assess it directly. Another issue not addressed in this study was whether the timing of the beliefs elicitation changes participants' guesses and their willingness to avoid their lying costs. Beliefs were elicited after P_1 has reported their random draw. Therefore, beliefs might be influenced by participants' reports. It was beyond the scope of the paper to test whether participants will be more prone to avoid their lying costs in AVOID when the elicitation task is done before reporting.

In spite of the mentioned limitations, the study certainly adds to our understanding of the role of prosociality on dishonesty. Although prosocial lying may seem trivial, it is, in fact, crucial in terms of today's concern over tax evasion and corruption. In practical terms, it suggests that having groups of people or intermediaries in positions where self-reports are central should be avoided because they will be more prone to lie. For instance, continued efforts are needed to make declaring taxes easier for the general population so that they do not need to use an intermediary to declare for them. The same principle applies to situations where one person is in charge of reporting the information on behalf of a team (for instance: a political party, a workgroup, or a firm). Individual reporting should always be preferred to creating dependencies between people's reports.

The findings provide important insights into the broader domain of dishonesty and prosociality. Nonetheless, some questions still remain to be answered. It is important to assess directly whether prosocial lying is a matter of intentions or consequences. In this paper, I found insights into consequentialism, but further research is needed to confirm it. Another natural progression of this work is to analyze whether in group settings reciprocal lying exists. Finally, further research might explore the role of the beliefs elicitation timing in strategic avoidance of lying.

References

Abeler, J., Nosenzo, D., and Raymond, C. (2019). Preferences for Truth-Telling. *Econometrica*, 87(4):1115–1153.

Akerlof, G. A. (1970). The market for “lemons”: Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3):488–500.

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The economic journal*, 100(401):464–477.

Andreoni, J. and Miller, J. (2002). Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica*, 70(2):737–753.

Ariely, D., Bracha, A., and Meier, S. (2009). Doing good or doing well? image motivation and monetary incentives in behaving prosocially. *American Economic Review*, 99(1):544–55.

Bénabou, R. and Tirole, J. (2006). Incentives and Prosocial Behavior. *The Amercian Economic Review*, 96(5):1652–1678.

Charness, G. and Rabin, M. (2002). Understanding social preferences with simple tests. *The quarterly journal of economics*, 117(3):817–869.

Chen, D. L., Schonger, M., and Wickens, C. (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Coffman, L. C. (2011). *Intermediation reduces punishment (and reward)*, volume 3.

Conrads, J., Irlenbusch, B., Rilke, R. M., and Walkowitz, G. (2013). Lying and team incentives. *Journal of Economic Psychology*, 34:1–7.

DellaVigna, S., List, J. A., and Malmendier, U. (2012). Testing for altruism and social pressure in charitable giving. *The quarterly journal of economics*, 127(1):1–56.

Dimant, E., Van Kleef, G. A., and Shalvi, S. (2020). Requiem for a nudge: Framing effects in nudging honesty. *Journal of Economic Behavior & Organization*, 172:247–266.

Dufwenberg, M. and Dufwenberg, M. A. (2018). Lies in disguise – A theoretical analysis of cheating. *Journal of Economic Theory*, 175:248–264.

Erat, S. and Gneezy, U. (2011). White lies. *Management Science*, 58(4):723–733.

Fries, T., Gneezy, U., Kajackaite, A., and Parra, D. (2021). Observability and lying. *Journal of Economic Behavior and Organization*, 189:132–149.

Garbarino, E., Slonim, R., and Villeval, M. C. (2019). Loss aversion and lying behavior. *Journal of Economic Behavior & Organization*, 158:379–393.

Gino, F., Ayal, S., and Ariely, D. (2013). Self-serving altruism? The lure of unethical actions that benefit others. *Journal of Economic Behavior and Organization*, 93:285–292.

Gneezy, U., Kajackaite, A., and Sobel, J. (2018). Lying Aversion and the Size of the Lie. *American Economic Review*, 108(2):419–453.

Hugh-Jones, D. (2019). True lies: Comment on garbarino, slonim and villeval (2018). *Journal of the Economic Science Association.*

Jiang, T. (2013). Cheating in mind games: The subtlety of rules matters. *Journal of Economic Behavior and Organization*, 93:328–336.

Kajackaite, A. and Gneezy, U. (2017). Incentives and cheating. *Games and Economic Behavior*, 102:433–444.

Karni, E. (2009). A Mechanism for Eliciting Probabilities. *Econometrica*, 77(2):603–606.

Khalmetski, K. and Sliwka, D. (2019). Disguising Lies—Image Concerns and Partial Lying in Cheating Games. *American Economic Journal: Microeconomics*, 11(4):79–110.

Kocher, M. G., Schudy, S., and Spantig, L. (2018). I lie? we lie! why? experimental evidence on a dishonesty shift in groups. *Management Science*, 64(9):3995–4008.

Leib, M. et al. (2021). (dis) honesty in individual and collaborative settings: A behavioral ethics approach.

Levine, E. E. and Schweitzer, M. E. (2015). Prosocial lies: When deception breeds trust. *Organizational Behavior and Human Decision Processes*, 126:88–106.

Mobius, M. M., Niederle, M., Niehaus, P., and Rosenblat, T. S. (2011). Managing self-confidence: Theory and experimental evidence. Technical report, National Bureau of Economic Research.

Muehlheusser, G., Roider, A., and Wallmeier, N. (2015). Gender differences in honesty: Groups versus individuals. *Economics Letters*, 128:25–29.

Palan, S. and Schitter, C. (2018). Prolific. ac—a subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17:22–27.

Potters, J. and Stoop, J. (2016). Do cheaters in the lab also cheat in the field? *European Economic Review*, 87:26–33.

Rilke, R. M., Danilov, A., Weisel, O., Shalvi, S., and Irlenbusch, B. (2021). When leading by example leads to less corrupt collaboration. *Journal of Economic Behavior Organization*, 188:288–306.

Shalvi, S. and De Dreu, C. K. (2014). Oxytocin promotes group-serving dishonesty. *Proceedings of the National Academy of Sciences*, 111(15):5503–5507.

Shalvi, S., Eldar, O., and Bereby-Meyer, Y. (2012). Honesty Requires Time (and Lack of Justifications). *Psychological Science*, 23(10):1264–1270.

Weisel, O. and Shalvi, S. (2015). The collaborative roots of corruption. *Proceedings of the National Academy of Sciences*, 112(34):10651–10656.

Wiltermuth, S. S. (2011). Cheating more when the spoils are split. *Organizational Behavior and Human Decision Processes*, 115(2):157–168.