

ARTICLE

Saliency, honesty and competition

Ceren Bengu Cibik¹ and Daniel Sgroi² 

¹Ofcom, London, UK and ²Department of Economics, University of Warwick, Coventry, UK

Corresponding author: Daniel Sgroi; Email: daniel.sgroi@warwick.ac.uk

This paper represents the views and opinions of the author(s) and should not be taken as a statement of Ofcom policy/opinion. Ethical approval reference ECONPGR 05/18. AEA RCT registrations AEARCTR-0005142 (wave 1) and AEARCTR-0005955 (wave 2).

(Received 11 March 2024; revised 6 May 2025; accepted 12 May 2025)

Abstract

We provide the first large-scale statistical investigation of the role of the saliency of (dis)honesty on future behavior in a multi-wave experiment with 1,260 subjects. In the first wave, we vary the saliency of subjects' past dishonesty and explore the impact on behavior in tasks that include the scope to lie. In the second wave, we vary the degree of competitiveness in one of our core tasks. In a real effort task with individual incentives, being asked to recall experiences that involve honesty, or dishonesty reduces dishonesty in the task. This effect persists, albeit with a smaller effect size, when we purposefully introduce competitive incentives to this task in wave 2. On the other hand, in a competitive environment in which subjects could earn more by lying to *their counterparts*, inducing them to think more about (dis)honesty pushes them toward becoming more dishonest.

Keywords: cognitive dissonance; competition; experiment; honesty; lying; salience

Introduction

We now have a sound grasp of how motivated reasoning and self-deception can be both ubiquitous and potentially very damaging for society as a whole (Benabou and Tirole 2016; Gino *et al.* 2016). What is less well understood is what happens when self-deception fails and individuals realize they are not the upstanding morally unambiguous individuals they might wish to be. In this paper, we provide an examination of what happens when an individual's own past experiences of (dis)honesty are made more prominent or salient.¹ Consider someone who is asked to recall a period of dishonesty in the past. Will this push them toward greater honesty or will this realization

¹The standard dictionary definition equates prominence with salience. Given the more common use of salience within the literature, we will also use this term. This is also related to the idea of 'what comes to mind' as discussed in Gennaioli and Shleifer (2010).

only increase their propensity for dishonesty in the future? The first, more positive, reaction to the increased prominence of dishonesty is normally referred to as *moral balancing*: an effort to correct the sins of the past by leaning toward greater morality in the future. The second, more negative response reflects the psychological costs associated with *cognitive dissonance*, the high cost of attempting to carry two opposing viewpoints at once, and might be expected to lead to acceptance and greater levels of dishonesty. Which force is more prevalent is an open and very topical question in a world increasingly characterized by dishonesty at all levels of society whether in our political and business leaders, our media (especially in an era of ‘fake news’) and in social media-based interactions in which lying can seem relatively easy, but where revelations about dishonest behavior in the past occur on a daily basis.

We run two waves of online experiment to answer this question. Wave 1 of our experiment involved 892 subjects. We first randomly assigned subjects to one of three treatment groups in which they were asked to write about incidents in their lives which involved dishonesty² of various types in the past. This is a priming technique developed in Fenigstein and Levine (1984) as a mean of making someone’s ‘ideal self’ more salient, activating self-related cognition and making the self more accessible as a causal agent. We then ask them to undertake two incentivized tasks, each of which included the scope to be dishonest. In one task, the ‘matrix puzzle’ subjects were asked to identify the number of cells in a matrix which sum to 10 and then report their answer earning higher payments for higher reported numbers.³ We compared both the absolute number reported in the treatment(s) vs the control as one measure of dishonesty but were also able to consider subjects who reported numbers higher than what was possible, and so must have lied. In another task, a cheap talk sender–receiver (deception) game taken from Gneezy (2005), we asked them to send a message to a partner, the message being either true or false with the likelihood being that the false message would result in higher rewards. In both tasks, they faced the incentivized temptation to lie though the contexts were different. We also collected responses to standard demographic questions as well as answers to various psychometric and personality questions. In wave 2, we repeated the basic setup with a further 368 subjects but asked subjects to undertake a single task: a variant of the matrix puzzle task where we modified the structure to raise the competitive nature of the task by making rewards partly dependent upon reporting a number that placed subjects inside the top 50% of the distribution. Our experiment enabled us to compare the treatment and the control groups to test whether becoming more aware of past (dis)honesty had an impact on behavior (which it did). We also compared the matrix puzzle with the sender–receiver game to assess whether competition and strategic interaction mattered (which they did), and finally, we compared the matrix task in wave 1 with the more competitive variant of the task in wave 2 to see if introducing competitive incentives had an effect (which it did).

We present results of the text analysis in Appendix A.

²Dishonesty in the context of this experiment refers to untruthful reporting or lying.

³This is a commonly used task, see Mazar *et al.* (2008). An example of a matrix task can be found in the Appendix.

Contribution to the related literature

Dishonesty, encompassing lying and cheating, has received considerable attention in academic research. Our study draws on this literature and utilizes two core tasks: and matrix puzzle (Mazar *et al.*, 2008; Gino *et al.*, 2009, 2013; Mead *et al.*, 2009; Shu *et al.*, 2011, 2011b; Kristal *et al.*, 2020)⁴ and a cheap talk sender–receiver game (Gneezy, 2005; Sutter, 2009; Erat and Gneezy, 2012; Gneezy *et al.*, 2014). Mazar *et al.* (2008) reported that participants, in the absence of cheating opportunities, achieved an average score of 3.4 matrices solved, which increased to 6.1 when cheating was possible. Notably, there was no significant variation in performance when the incentive for cheating was raised to \$2 per correct answer.⁵ Gneezy (2005) and Erat and Gneezy (2012) investigated the consequences of lying on dishonest behavior by using a sender–receiver game.

Both papers classified the lie depending on the consequence of the lie on the liar and other people to differentiate the incentives behind one's decision to behave dishonestly.⁶ Erat and Gneezy (2012) found that the percentage of senders who lied varied depending on the cost and benefit involved. Specifically, participants were more reluctant to tell a Pareto white lie, which benefits both the liar and others, compared to an altruistic white lie, which benefits others at the expense of the liar. Similarly, Kerschbamer *et al.* (2019) showed in their experiment that altruistic individuals lie less when their lies harm others but do not lie less when lying benefits them without affecting others. These results add another layer to our understanding of dishonesty, indicating that the motivations behind dishonest behavior can be multifaceted and context-dependent.

To the best of our knowledge, no study has directly compared the sender–receiver game with any of the other experimental paradigms, resulting in different conclusions across these paradigms (Gerlach *et al.*, 2019). In our study, we are utilizing these two distinct measures of dishonesty that differ in terms of the decision environment to study the complexity of dishonesty in social contexts. This approach enables us to investigate the impact of strategic behavior and competitive incentives on dishonest behavior specifically when (dis)honesty is made salient. Our study makes a significant and novel contribution by delving into this unexplored interaction between saliency, (dis)honesty and competitive incentives, shedding light on a crucial dimension that has been largely disregarded in previous research.⁷

⁴Despite replication challenges (Verschuere *et al.*, 2018), the matrix puzzle remains widely utilized and its effectiveness as a research method is not dependent on this debate.

⁵Another popular experimental tool used to measure dishonesty is a simple coin flip or a rolling-a-die task where participants are paid on the basis of the number of heads (or tails) reported (Buccioli and Piovesan, 2011; Houser *et al.*, 2012; Fischbacher and Föllmi-heusi, 2013; Abeler *et al.*, 2014; Cohn *et al.*, 2015; Dimant *et al.*, 2020). For instance, Abeler *et al.* (2014) found that 55% of the subjects reported an outcome that provided no material payoff even though there was no incentive to behave honestly in their experiment. Field experiments are also widely used in the dishonesty literature, for instance, Stoop (2014) (lost/misdirected letters), Cohn *et al.* (2019) (wallets found on the street where the return rate of the lost items determines the level of honesty among the sample) and Pruckner and Sausgruber (2013) (newspaper sales on the street). See Rosenbaum *et al.* (2014) and Gerlach *et al.* (2019) for meta-analysis on dishonesty.

⁶We designed our treatments based on this classification of lies.

⁷Previous studies such as Faravelli *et al.* (2015), Aydogan *et al.* (2017) and several others referenced in Gerlach *et al.* (2019)'s meta-analysis have solely examined the relationship between competition and

Priming techniques have also been widely used in the economics and psychology literature while studying dishonesty and immoral behavior, especially in the form of honor, religious, moral and legal codes (Dickerson *et al.*, 1992; McCabe and Trevino, 1993, 1997; Mazar *et al.*, 2008; Shu *et al.*, 2011a; Pruckner and Sausgruber, 2013; Aveyard, 2014; Kettle *et al.*, 2017). Diener and Wallbom (1976) provide the first method to generate self-awareness by making people sit in front of a mirror and listen to their self-recorded tape about their physical characteristics and occupation.⁸ We adopted Fenigstein and Levine (1984)'s priming technique which is a story construction task to directly access self-referent thoughts. They hypothesized that writing a personal narrative prompts individuals to assess their causal effect on outcomes, activating self-related cognition and increasing salience of the self.⁹ Our study's main divergence from the aforementioned studies and key contribution is that we are the first to examine the effect of salience on dishonesty within distinct choice environments characterized by varying levels of strategic and competitive incentives. We clarify why the different experimental paradigms yield different conclusions regarding dishonesty as discussed in Rosenbaum *et al.* (2014) and Gerlach *et al.* (2019).

Experimental design

The online experiment itself consisted of three stages and two waves. The first stage included a questionnaire containing basic demographic questions, together with questions designed to elicit preferences for risk, integrity, ethical stance, fairness (the Ultimatum Game) and a brief version of the Big Five Inventory designed to detect personality traits (Rammstedt and John, 2007). In the second stage of the experiment, subjects were randomly assigned to one of four different groups: control group, honesty treatment, low dishonesty treatment, or high dishonesty treatment. In these three treatment groups, subjects were asked to write about a real-life event that took place in the 12 months before the experiment, during which they were completely honest (the honesty treatment), dishonest but with no negative effect on others (the low dishonesty treatment) or dishonest with a negative effect on others (the high dishonesty treatment).¹⁰ In the control group, subjects were not asked to write anything but instead progressed directly to the third stage of the experiment. We made a conscious decision to use a no-intervention (passive) control group to overcome various issues related to confounding.¹¹ This stage aimed to generate between-subject variation in the

immoral behavior. However, these studies have not addressed the importance of the saliency of (dis)honesty while investigating this relationship.

⁸See Basic and Quercia (2020), Bender *et al.* (2018) and Falk (2021) as examples of their method employed in other experiments.

⁹See Jordan *et al.* (2011) as another example of use of this priming method.

¹⁰It is important for us to differentiate the incentives behind one's decision to behave dishonestly in order to fully capture the heterogeneity in the decision-making process. As a result, we base our treatments on the classification of lies by (Gneezy, 2005; Erat and Gneezy, 2012). Even though the psychological cost of lying might be different between spiteful black lies and selfish black lies, we expect all treatment groups to make dishonesty salient.

¹¹This is discussed further in Appendix A.

saliency of dishonesty and honesty. The last stage of the experiment contained incentivized tasks, undertaken in a random order, which included the option to be dishonest which we call 'dishonesty tasks' for simplicity. Subjects were paid a bonus based on their actions in one of these dishonesty tasks chosen at random at the end of the experiment. The two waves differed only concerning the dishonesty tasks used.

The first wave of the experiment included two dishonesty tasks: the matrix puzzle game (Mazar *et al.*, 2008) and the cheap talk sender–receiver game (also called the 'deception game' by Gneezy (2005)).¹² In the matrix game, participants were presented with an image containing 20 different matrices and asked to find as many pairs of numbers that sum to 10 in these matrices as possible within 5 minutes. Once the time was over, subjects were directed to the next page to report the number of pairs they found. The potential bonus they made depended on the number of matrices they reported solving. The experiment included two variations where the incentives were \$0.10 (low incentive) and \$0.30 (high incentive) per correct answer.¹³ Since the subjects' answers could not be confirmed, they were free to report any number between 0 and 20, and it was not possible to detect dishonesty at the individual level *unless* they reported an infeasible large number.¹⁴ The number of reported answers is normally assumed to be increasing in the level of dishonesty of the subject. As a special property of our design, the maximum number of correct answers was 10, and therefore any reported number greater than 10 was considered to be infeasible. Both tasks (and the different incentive versions of the tasks) were presented in a randomized order.

In the sender–receiver game, subjects were told that they were matched with another anonymous MTurk worker and played the role of the sender. We matched senders with a randomly chosen receiver from an earlier online experiment that employed the same game with identical payoffs to decide their bonus payments in case this task was randomly selected as the bonus payment task. The other worker (the receiver) made the final decision about which of two possible options, 'Option A' or 'Option B', would be selected, which determined the payoffs of both players. In our design, two different allocations related to Option A and Option B were used but in both of these two cases, Option B always gave a higher payoff to the sender (respectively a lower payoff to the receiver) than Option A. However, the payoffs associated with these two options were visible only to the sender, not to the receiver. Instead, the subject, who did see the payoffs associated with the two options, had the opportunity to send a message to the receiver to help guide their choice.¹⁵ In our design, the payoff allocations were set as \$1

¹² Both tasks (and the different incentive versions of the tasks) were presented in a randomized order.

¹³ Participants completed both versions of the same task in a randomized order.

¹⁴ We did not explicitly make subjects aware of whether and how their dishonesty would be detected at the individual level, as was done in the original version of the task.

¹⁵ Sutter (2009) argues that telling the truth should also be considered a form of deception if the main purpose of sending a truthful message is expecting that the other person will not follow his message. He calls this phenomenon 'sophisticated deception' and we might also consider it to be second-order dishonesty since it relies on the actions of others. This is not something we specifically address in our study; however, the most relevant prior literature suggests that this is unlikely to be an issue. For instance, Gneezy (2005) find that 82% of the senders reported that they expected the receiver to follow their message and 78% of the receivers did indeed follow the message. In addition, Dreber and Johannesson (2008) and Sutter (2009) showed that the majority of receivers choose to follow the message sent by the sender and most senders expect this. Even

for the sender (subject) and $\$X$ for the receiver under option A and $\$X$ for the sender (subject) and $\$1$ for the receiver under option B, where X was set equal to $\$1.20$ in the low incentive setting, and $\$3$ in the high incentive setting.¹⁶ The subject was given two possible messages to send: 'Option B will earn you more money than Option A' (which was not true and was hence classified as a dishonest message) or 'Option A will earn you more money than Option B' (which was true and was hence classified as an honest message).

The second wave of the experiment differed from the first wave in only one respect: it included only one dishonesty task, a modified version of the matrix puzzle game. As in the wave 1 version of the task, subjects were asked to report the number of pairs that they found in the matrix puzzle which add up to 10. The only difference from the wave 1 task was the payment scheme. While in wave 1, participants were paid for each matrix they reportedly solved, in the modified version, they were paid a lump-sum amount only if they were in the top 50% of the distribution. The lump-sum amount was decided based on the difference between the average payment that participants in the top 50% of the distribution received and the average payment that participants in the bottom 50% received in wave 1 of the experiment. This added competitive incentives to the matrix puzzle task, but maintained the same average payment in expectation by applying a mean-preserving spread by awarding a bonus only to those in the top half of the distribution.¹⁷ As in the wave 1 task, the experiment included two versions of this task with low or high material incentives -played in a randomized order- allowing us to check whether our results were responsive to a change in the incentive payments. The lump-sum payments were set to be $\$0.72$ in the low incentive task and $\$2.13$ in the high incentive task and were calculated to yield the same average payment as in the wave 1 variant of the game.

The dishonesty tasks were followed by two dictator games designed to elicit subjects' preferences for altruism. They were asked to indicate what percentage of their actual bonus from this experiment they would like to donate to Macmillan Cancer Support and/or to the researchers of this experiment to be used for research purposes. By employing dictator games, we aimed to test whether the moral balancing argument held in our sample. These games gave subjects the opportunity to balance their earlier dishonesty by donating either to charity or to researchers. Therefore we could directly compare actions in the dishonesty tasks with donation levels. The last task in

though we did not elicit senders' beliefs in this experiment, we would expect to observe the same result in our sample. Regarding receivers' beliefs, we have unreported results from a pilot in which we found that 78% of receivers followed the message sent by the senders. In other words, hoping for receivers to do the opposite from the sent message would not be a winning strategy. Regarding our data, Table A1 indicates that the strategic play could not be the main motivation for senders as participants in the treatment group donated significantly less to the charity and the researcher as compared to control group, showing the negative effect of salience of (dis)honesty on altruism. However, we are not able to differentiate different motives clearly with our experiment.

¹⁶Subjects completed both high and low incentive variations of this game in a randomized order.

¹⁷We categorize our tasks across four characteristics as discussed in the 'Main hypotheses' section. By adding a competitive element to the matrix puzzle, we are manipulating one of these variables to observe its effect on dishonesty.

the final stage of the experiment aimed to provide data to enable us to check for possible experimenter demand effects.¹⁸

The experiment was conducted in 2020 on Amazon M-Turk, the first wave in February and the second wave in July.¹⁹ Subjects earned a show-up fee of \$2 plus a performance related bonus payment (\$0.42 on average). The experiment took approximately 25 minutes on average. Out of the 892 subjects who took part in wave 1, 284 were randomly allocated to the control group, 205 to the honesty treatment group, 208 to the low dishonesty treatment group and 195 to the high dishonesty treatment group. Out of the 368 subjects who took part in wave 2, 101 were randomly allocated to the control group, 76 to the honesty treatment group, 104 to the low dishonesty treatment group and 87 to the high dishonesty treatments group.^{20,21} Appendix A provides an ex-post verification that the filtering rule was sensible based on text analysis performed on the text written by participants in the experiment. A simplified timeline of events is presented in Table 1. Full experimental instructions can be found in Appendix B.

Main hypotheses

In this section, we state six testable hypotheses that are formed by relevant discussions in the existing literature.

According to the rational choice model by Rabin (1994), there are three important variables that affect people's choice to behave dishonestly. They are the material benefit obtained from dishonesty, the psychological cost of dishonesty (i.e. cognitive dissonance), and the cost of developing beliefs that are not consistent with the honest set of beliefs about the morality of dishonesty.^{22,23} He argues that immoral behavior generates

¹⁸We employed three attention questions throughout the experiment to make sure that participants were actually following the instructions. These questions along with the experimental instructions can be found in Appendix B.

¹⁹Note that in most cases in the results to follow, we will compare control and treatment groups within waves and so the temporal gap between waves is not relevant. In the cases where we do compare between waves, we will always include a full set of demographic variables to control for any demographic changes that might have taken place between the two waves.

²⁰Initially, we collected 1110 observations in wave 1 and 705 in wave 2. Before starting to analyze our data, we applied a relevance filter on the text. The filter is listed in our preregistration and is also validated by the text analysis in Appendix A. In total, we eliminated 218 subjects in wave 1 and 335 subjects in wave 2 who wrote an irrelevant text in the second stage of the experiment.

²¹Even though the exclusion rate based on the relevance filter is higher in wave 2, we do not observe any systematic difference in the exclusion rate among different treatment groups. The exclusion rate in wave 1 is 26.52% in the honesty treatment, 24.64% in the low dishonesty treatment and 28.04% in the high dishonesty treatment. The exclusion rate in wave 2 is 61.02% in the honesty treatment, 49.52% in the low dishonesty treatment and 56.72% in the high dishonesty treatment. See Appendix A for further analysis on the relevance filter.

²²His model suggests that if a person receives lower material utility from engaging in an immoral activity, or it becomes more costly to maintain modified beliefs about the morality of behavior, or there is greater distaste for cognitive dissonance, then that individual would reduce the amount of the perceived immoral activity.

²³We consider the cost of developing beliefs that are inconsistent with one's honest set of beliefs about dishonesty under the psychological cost of dishonesty as it is not possible to distinguish them in our experiment.

Table 1. Experimental design

	Control group	Honesty treatment
First Stage	Demographic Questionnaire, Ethic and Integrity Questionnaire ^a , Risk preference ^b , Ultimatum Game, the Big Five	
Second Stage	Writing about an event in their own life which involves complete honesty	
Third Stage	Dishonesty Tasks ^c , Dictator Game, Demand Effect questions	
	Low Dishonesty Tr.	High Dishonesty Tr.
First Stage	Demographic Questionnaire, Ethic and Integrity Questionnaire ^a , Risk preference ^b , Ultimatum Game, the Big Five	
Second Stage	Writing about an event in their own lives in which they decided not to be completely honest to themselves but it did not end up harming someone else	Writing about an event in their own lives in which benefit they decided not to be completely honest to benefit themselves but it ended up harming someone else.
Third Stage	Dishonesty Tasks ^c , Dictator Game, Demand Effect questions	

Notes: ^aThe integrity questionnaire is taken from Whiteley (2012). ^bTo elicit preferences toward risk, we follow the method outlined in Gneezy and Potters (1997). ^cDishonesty tasks include the Matrix Puzzle and the Cheap Talk Sender–Receiver Game in wave 1 and the modified Matrix Puzzle in wave 2. Subjects were paid a bonus based on their actions in one of these dishonesty tasks chosen at random at the end of the experiment. Both tasks (and the different incentive versions of the tasks) were presented in a randomized order.

a cost that arises through the existence of cognitive dissonance. This can be described as the conflict between the benefits of believing in one’s ideal self and the costly realization that the true self does not meet this standard. In our experiment, we vary material incentives and the psychological cost of dishonesty and focus specifically on two main factors that might affect the level of cognitive dissonance experienced: the saliency of (dis)honesty and the context where the decision is made.

First, we ask whether the saliency of (dis)honesty matters. We expect that writing about past dishonesty will impose psychological discomfort due to the conflict between their ideal self and their true self, and this cognitive dissonance will affect future dishonest behavior.²⁴

Hypothesis 1: The saliency of (dis)honesty affects dishonest behavior.

Our next two hypotheses address the issue of context.

Comparison of different dishonesty tasks

To explore the effect of context in which decision is made, we categorize our tasks based on four characteristics. These characteristics help determine the level of cognitive dissonance or psychological cost that comes about through dishonest behavior. They are (i) the degree of competition in the game, (ii) whether dishonesty is salient,

²⁴We investigate the effect of priming self-identity on dishonesty with a two-sided hypothesis given the conflict between ‘moral balancing’ and ‘moral reinforcement’ discussions in the literature (Merritt *et al.*, 2010; Mullen and Monin, 2016). See the ‘Discussion’ section for further discussion on this conflict.

(iii) whether the final decision is only the responsibility of one person and (iv) whether the task is ego-relevant or not. The impact and even direction of the effect of the context on the level of dishonesty would depend on these characteristics which help us to characterize our next two hypotheses (Hypothesis 2 and 3).

Firstly, behavior that might seem to be unambiguously dishonest in a task with individual incentives, might seem more acceptable if competitive incentives were involved when 'defeating' your rival is the ultimate aim and so damaging the other player's payoff is a key component of the game, bringing to mind the proverb 'all is fair in love and war'. Research on moral hypocrisy (Valdesolo and DeSteno, 2007, 2008) suggests that even when participants do behave badly (dishonestly) in a game setting by taking advantage of experimental partners, they are motivated to reinterpret their own behavior as less bad in order to maintain their moral self-identity. Additionally, in situations where lying is explicitly included as one of the available actions in a game, (or when awareness of cheating is heightened as demonstrated in Fosgaard *et al.* (2013), or when there is an increased tendency to conform to cheating as observed in Kocher *et al.* (2018), individuals may find it easier to mentally categorize such actions as selecting a strategy provided by the experimenter rather than engaging in dishonest behavior. This perception creates a form of moral 'wiggle room' that helps reduce the salience and perceived cost associated with lying, thus aiding in overcoming cognitive dissonance. Thirdly, if the payoffs in the game depend not only on the player's action itself but also on the other player's action, it might become easier to feel a lower level of responsibility for any dishonest action. Consider, for instance, the ultimate decision of whether to believe and act upon a lie is in the hands of another person (as in the case of the sender–receiver game in which the sender may send a dishonest message, but the receiver chooses whether to act upon this lie). Finally, if a game is ego-relevant, the cost of lying might be higher since it includes elements of self-deception or the formation of motivated beliefs which come with a cost (as in Benabou and Tirole (2016)), as well as the deception of others.

As we argued above in our categorization of the tasks if dishonest behavior in a task is expected to create a higher (lower) level of cognitive dissonance, regardless of the past positive or negative self-primed, the level of dishonesty will be lower (higher) in this task than in a task where the dishonest behavior is associated with lower (higher) levels of cognitive dissonance. We can refine this logic further so that it applies directly to the specific form of games, which are described in detail in the 'Experimental design' section. We expect that reporting dishonestly in the matrix puzzle game will cause experiencing a higher level of cognitive dissonance than in the sender–receiver game. This is because the matrix puzzle is a single-player, real-effort task without competitive or strategic incentives, is ego-relevant, and does not present untruthful reporting as an option given by the experimenter. Conversely, the sender–receiver game, where deception is less salient and responsibility is shared, allows for greater dishonest behavior. This discussion helps us characterize our next hypothesis:

Hypothesis 2: Context matters: The saliency of dishonesty stemming from any of our treatments should result in a reduction in dishonesty levels in the matrix puzzle game, but an increase in dishonesty levels in the sender–receiver game.

Based on the discussion above, we also expect that reporting dishonestly in the modified matrix puzzle game (wave 2) will cause lower cognitive dissonance than in the original matrix puzzle game (wave 1). This is because the modified game includes a strategic component that could cause negative externalities on others for personal material gain. Therefore, they should behave more dishonestly in wave 2 version of this game. This discussion yields our next hypothesis.

Hypothesis 3: Context matters: The saliency of dishonesty stemming from any of our treatments should lead to a greater reduction in dishonesty levels in original matrix puzzle game (wave 1) compared to the modified matrix puzzle game (wave 2).

As a related supplementary hypothesis, we can also consider the extent to which material incentives influence behavior and change the nature of the relationship between salience and dishonesty (Kajackaite and Gneezy, 2017; Gneezy *et al.*, 2018). To that end, we have our next testable hypothesis, which suggests that material incentives may play a role in determining the relationship between salience and dishonesty.

Hypothesis 4: Material incentives plays a role in determining the relationship between saliency and dishonesty.

As an additional exercise, we might also consider the characteristics of those who lie in a regular and detectable way. We will do so in the ‘Results’ section to follow and can here identify two testable hypotheses relating to consistency and moral balancing:

Hypothesis 5: (Consistency) Those who lie more in one task are likely to lie more in the other.

Hypothesis 6: (Moral balancing) Lying more should result in higher donations to charity and/or the researcher.

Finally, with the availability of text data from the main treatment, we are in a position to carry out text analysis in an attempt to look for significant differences in sentiment and style between the text written in the different treatments. This will necessarily be exploratory but should also serve to validate the effectiveness of the treatment.

Results

In what follows, we will structure our results first into an attempt to investigate the core hypotheses relating to the link between salience, context and dishonesty (identified above as Hypotheses 1–4) via comparisons between the treatment and control groups before moving on to a discussion of the characteristics of lying that are detectable at the individual level (relating to Hypotheses 5 and 6). Note that our results are designed to link easily to the corresponding hypothesis, so Result 1 refers to Hypothesis 1, and so on.

Table 2. Mean value comparisons of various dishonesty tasks

	Wave 1				Wave 2	
	Matrix Puzzle		CT Sender–Receiver Game		Modified Matrix Puzzle	
	No of matrix reported to be solved		% of people who sent a dishonest message		No of matrix reported to be solved	
	Low incentive	High incentive	Low incentive	High incentive	Low incentive	High incentive
Mean Values						
Control Group	5.746	5.799	0.394	0.500	7.881	7.059
Honesty Treatment	4.766	4.868	0.576	0.634	5.408	6.053
Low Dishonesty Tr.	4.822	4.827	0.514	0.543	5.712	5.356
High Dishonesty Tr.	5.528	5.487	0.497	0.595	5.138	4.885
Treatment Groups	5.030	5.050	0.530	0.590	5.438	5.401
t-test^a						
Honesty vs Low Dishonesty	0.908	0.930	0.213	0.061*	0.696	0.355
Honesty High Dishonesty	0.130	0.225	0.118	0.421	0.714	0.120
Low Dishonesty vs High Dishonesty	0.172	0.197	0.734	0.297	0.426	0.503
Control vs Treatment	0.053*	0.042*	0.0002*	0.0112**	0.0001***	0.0055***

Notes: ^a p-Values from a two-tailed t-test are reported. *p < 0.10, **p < 0.05, ***p < 0.01.

Group-level results

We can first start with a mean value comparison test between the treatment groups and the control group for the dishonesty variables obtained from the different dishonesty tasks, which will provide us with one very useful insight. Table 2 reports the mean values for all dishonesty tasks conducted in the first and second waves of the experiment for the treatment groups and the control group separately. The findings in this table allow us to reveal our only result that did not form part of our preregistered hypotheses: different types of treatment groups seem to behave in a similar way to each other when compared with the control group.

Result 0: The focus on honesty or dishonesty in the treatment does not seem to matter when compared with behavior in the control.

In other words, what seems to matter is making (dis)honesty more salient rather than the way in which this salience comes about. Recalling a recent event where the person was honest might also bring up stories about dishonesty and vice versa. This result allows us to introduce a simplification into the reporting of our findings enabling us to combine the treatment groups to conduct our analysis. This allows us to undertake a more parsimonious (single) treatment vs control comparison, and also helps minimize the need for p-value adjustments from multiple hypothesis testing. This result might also be useful for future experimental designs in this area since it suggests that experimental researchers need only chose to focus on sparking recollections of honesty or dishonesty in order to make honesty salient.

Asking subjects to recall recent experiences with honesty or dishonesty has a significant impact on their future behavior in the experiment. We can see from Table 2, the treatment group significantly differs from the control group for all dishonesty variables. This result provides immediate support for Hypothesis 1 which confirms the importance of the saliency of dishonesty.

Result 1: Saliency matters. Changes in cognitive dissonance affect the level of dishonesty in the future. Moreover, this impact is largely neutral to the type of treatment.

To investigate whether the direction of the effect is determined by the context where people make their decision, we compare the effect of the saliency of dishonesty, differentiating by dishonesty tasks. In line with Hypothesis 2, saliency decreases the level of dishonesty in the matrix puzzle game, whereas it increases the level of dishonesty in the sender–receiver game in wave 1. Table 2 shows that the mean number of matrices reportedly solved by subjects in the control group is 5.746 (5.799), whereas it is 5.03 (5.05) in the treatment group for the low (high) incentive task. For subjects playing the matrix puzzle game, saliency significantly lowers the level of dishonesty. On the other hand, the proportion of people who sent a dishonest message in the control group in the low (high) incentive sender–receiver game is 39.4% (50%), whereas it is 53% (59%) in the treatment group. We see a significant increase in dishonesty after (dis)honesty has become more salient in the sender–receiver game.

Table 3. Regression analysis

Variables	Matrix Puzzle		Sender–Receiver Game	
	(1)	(2)	(3)	(4)
	Model 1	Model 2	Model 1	Model 2
Treatment	–2.051*** [0.601]	–1.932*** [0.609]	0.113*** [0.0304]	0.108*** [0.0306]
Wave 1	–1.840*** [0.626]	–1.558** [0.628]	–	–
Treatment × Wave 1	1.319* [0.700]	1.279* [0.703]	–	–
High Incentive	–0.253 [0.196]	–0.256 [0.197]	0.0758*** [0.0172]	0.0773*** [0.0174]
High Incentive × Wave 1	0.285 [0.223]	0.293 [0.224]	–	–
Constant	7.597*** [0.545]	8.003*** [1.641]	–	–
Observations	2,520	2,512	1,784	1,778
R-squared	0.016	0.040	0.0121	0.0252
Control Variables	✗	✓	✗	✓

Notes: Standard errors are clustered at the individual level and shown in brackets. Columns 1 and 2 represent the linear regression results on the number of matrices reported to be solved in the matrix puzzle game. Columns 3 and 4 represent the probit regression results on a variable which takes 1 if a dishonest message is sent in the sender–receiver game and 0 if an honest message is sent. Control variables include age, age square, being married, having at least a college degree and being American. Stars indicate statistical significance as follows: * $p < 0.10$, ** $p < 0.05$ and *** $p < 0.01$.

Result 2: Context matters: The saliency of dishonesty leads to a decrease in dishonesty in the matrix puzzle game but also lead to an increase in dishonesty in the sender–receiver game.

In order to test Hypothesis 3, we conducted a second wave of the experiment in which the matrix puzzle game was modified in an attempt to lower expected cognitive dissonance after dishonest behavior, as compared with the wave 1 version. We added a strategic component to the task by introducing a competitive incentive setting. We expect to observe higher levels of dishonesty in the wave 2 version of the matrix puzzle game than in the wave 1 version. The last two columns of Table 2 show that the mean number of matrices reported to be solved in the modified matrix puzzle game is 7.881 (7.059) in the control group, whereas it is 5.438 (5.401) in the treatment group for the low (high) incentive task. This result shows a significant decrease in dishonesty after the past (dis)honest self made more salient and an increase in overall dishonesty when it is compared with the original version of the matrix puzzle game.

To delve deeper into this effect, we present a regression analysis in Table 3 which supports the results obtained from the mean value comparison tests. The first two columns of the table represent linear regression results on the number of matrices reportedly solved in wave 1 and wave 2. The last two columns present the marginal

effects from the probit regression on a dummy variable which takes the value of 1 if a subject sent a dishonest message in the sender–receiver game in wave 1. Model 1 includes only the main variables for both of the regressions, whereas model 2 adds demographic control variables that were collected at the beginning of the experiment. In both models, we merged the data from waves 1 and 2 and since only wave 1 features a sender–receiver game, the last two regressions do not include wave 1 variables or interactions. ‘Wave 1’ is a dummy variable which takes the value of 1 if the observation is drawn from wave 1 of the experiment and 0 if it is drawn from wave 2. The ‘treatment’ variable takes the value of 1 if the observation belongs to the treatment groups and 0 if it belongs to the control group. We treat the low-incentive and high-incentive tasks as one and include a dummy variable ‘incentive’ which takes the value of 1 if the observation is from the high-incentive version of the game and 0 if it is from the low-incentive version. The regressions also include an interaction of the treatment variable with wave 1 to observe whether the effect of the treatment differs among the two different versions of the matrix puzzle game across waves and an interaction of the incentive variable with wave 1 to observe whether the effect of material incentives differs among waves.²⁵

Column 1 of Table 3 suggests that subjects in the treatment group report 2.051 fewer total matrices than subjects in the control group ($p < 0.01$), which is in line with the Result 2. We also note an increase in the number of matrices reported in wave 2 as compared to wave 1. Subjects in wave 2 report 1.84 additional matrices on average than subjects in wave 1 ($p < 0.01$). This result supports Hypothesis 3 which states that since the induced level of cognitive dissonance (the psychological cost of lying) is lower in the wave 2 version of the matrix game, the level of dishonesty should be higher. The interaction of the treatment and wave variables is also significant at the 10% level. Combining the three significant variables in our analysis suggests that the level of dishonesty is significantly higher in the control group than the treatment group by 0.731 units in wave 1 and by 2.051 units in wave 2. Moreover, dishonesty is significantly higher in wave 2 than in wave 1 by 0.621 units in the treatment groups and by 1.84 units in the control group. Column 2 adds control variables to model 1: subjects in the treatment group report 1.932 fewer matrices than subjects in the control group ($p < 0.01$) and subjects in wave 2 report 1.558 more matrices solved than subjects in wave 1 ($p < 0.05$). The interaction of the treatment with the wave variable remains significant at the 1% level which suggests that the level of dishonesty is significantly higher in the control group than the treatment group by 0.653 units in wave 1 and by 1.932 units in wave 2. Dishonesty is significantly higher in wave 2 than in wave 1 by 0.279 units in the treatment groups and by 1.558 in the control group. Taken together this generates our next result which indicates a higher level of dishonesty in the matrix puzzle game in wave 2 than in wave 1:

²⁵In an unreported regression, we added the interaction between the treatment variable with the incentive variable to the models in Table 3. However, since the variable was not significant ($p > 0.10$) in any of the regressions, we did not include it in this table.

Result 3: Context matters: The level of dishonesty is higher in wave 2 than wave 1 for both control and treatment groups, as cognitive dissonance stemming from a dishonest behavior is lower in wave 2 version of the game.²⁶

Our design allows us to detect lying at the individual level if subjects report an infeasible number of solved matrices. In particular, some subjects reportedly solved more than 10 matrices in both versions of the game which is not possible since there were only 10 matrices that included two numbers which summed to 10. While this is perhaps not as robust a distinction as in our more general results above (because individuals can lie while remaining within the bounds of feasibility), we can see some further support for our findings coming from this quarter, for example, our data reveal that 13.6% of the wave 1 control group (of 284 subjects) reportedly solved more than 10 matrices, with this number rising to 20.8% for those in wave 2 which are different at the 5% level of significance ($p < 0.028$). We make further use of this feature of the experiment later when we try to identify key characteristics of liars which require lying to be detectable at the individual level. We next move on to examining the role of additional material incentives in decision-making. The incentive variable and the interaction of incentive with the wave variable are not significant for the matrix puzzle game in any of the models. This suggests that increasing material benefits does not affect the level of dishonesty which is in line with the results in Mazar *et al.* (2008). Column 3 presents the marginal effects from a probit regression on the decision to send a dishonest or honest message in the sender–receiver game in wave 1. In both model 1 and model 2, the treatment effect is significant and positive suggesting that subjects in the treatment group are more likely to send a dishonest message than subjects in the control group ($p < 0.001$). As stated in Hypothesis 2, we observe an increase in dishonesty when subjects' past (dis)honest behavior is made more salient. This is in line with our earlier argument that the sender–receiver game incorporates a relatively low psychological cost of lying attributable to cognitive dissonance. Moreover, as opposed to the results from the matrix puzzle game, material benefit is a significant determinant of whether to send a dishonest or honest message in this game. The results suggest that there is an increase in the probability of sending a dishonest message by 7.58% (7.73%) in the sample when the material benefit of lying is increased from \$0.2 to \$2.0. This result is consistent with the findings in Gneezy (2005). This leads us to our next result:

Result 4: Material incentives do not play a significant role in behavior in the matrix puzzle game but do play a significant role in the sender–receiver game.

²⁶While this may partially be attributed to lower cognitive dissonance stemming from dishonest behavior in the wave 2 version of the game, the competitive environment introduced in wave 2 also raises the material gains associated with lying. The increase in material benefits could simultaneously make dishonesty more costly. This suggests that differences in dishonesty levels could reflect both psychological costs and incentives tied to material benefits. While we acknowledge that the added element of competitive incentives in wave 2 increases the expected material benefits of dishonesty, we note that this experiment does not allow us to differentiate between the effects of cognitive dissonance and material incentives.

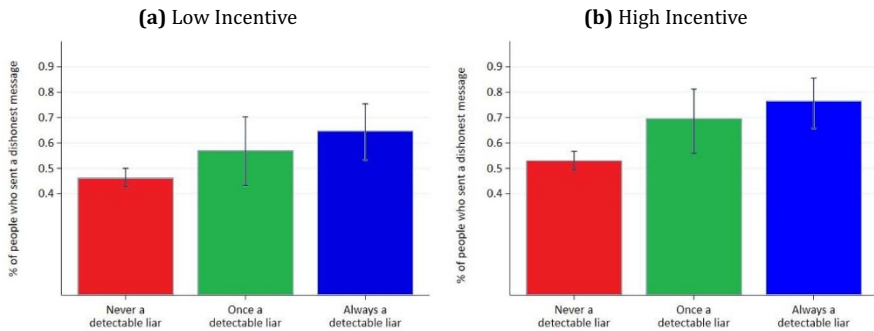


Figure 1. Comparison of liars across dishonesty tasks. (a) Low incentive. (b) High incentive.

The figure represents the proportions of people who sent a dishonest message in the sender–receiver game for people who were never, once, or always detectable liars (reported more than 10 matrices to be solved) in the matrix game (includes both low and high incentive games). Panel A corresponds to the low-incentive sender–receiver game, and Panel B corresponds to the high-incentive sender–receiver game.

Individual-level results

Next, we compare the detectable incidence of lying across tasks. We label someone as a ‘detectable liar’ if they reported more than 10 solved matrices in the matrix puzzle game: something that cannot possibly be true. While most of the results above are based on group comparisons since we cannot generally know how many matrices an individual truly solved, this classification gives us access to individual-level data and allows us to consider the characteristics of those who report more than 10 solved matrices. Figure 1 classifies the subjects as ‘never a detectable liar’ if they did not report more than 10 solved matrices in any of the low or high incentive matrix puzzles in wave 1, ‘once a detectable liar’ if they reported more than 10 matrices in only one of the matrix puzzles and ‘always a detectable liar’ if they reported more than 10 matrices in both low and high incentive matrix tasks.²⁷ Note that we cannot rule out that individuals lied even if they reported less than 10 solved matrices and that is the basis of the group-level comparisons above which instead considers the average number solved.

Figure 1 shows that the proportion of people who sent a dishonest message in the sender–receiver game increases with the frequency of reporting more than 10 matrices in the matrix puzzle game. Moreover, 46.38% (53.10%) of subjects who are never detectable liars in the matrix puzzle games sent a dishonest message in the low (high) incentive sender–receiver game, whereas 57.14% (69.64%) of subjects who were once detectable liars sent a dishonest message in the low (high) incentive sender–receiver game and that percentage shifts to 64.94% (76.62%) for those who are always a detectable liar. For the low incentive task, the confidence intervals and a two-sided *t*-test indicate a significant increase in the proportion of people who sent a dishonest message in the sender–receiver game from those who are never detectable liars to those who are always detectable liars ($p < 0.01$).

²⁷Note that this is the only classification possible for both the matrix puzzle and the sender–receiver game as there were only two incentive tasks in the experiment.

Table 4. Who are the liars?

	Sender–receiver game		
	(1) Never Lied	(2) Lied Once	(3) Always Lied
<i>Dishonesty Variables</i>			
No of matrix (Low Inc.)	4.583	5.183	5.9***
Detectable Liars % (Low Inc.)	7.82	9.36	16.86***
No of matrix (High Inc.)	4.547	5.306*	5.931***
Detectable Liars % (High Inc.)	5.54	11.49**	17.43***
<i>Personality Variables</i>			
BFI-Extraversion	0.261	0.296**	0.281
BFI-Conscientiousness	0.488	0.488	0.512**
BFI-Neuroticism	0.284	0.276	0.258*
BFI-Agreeableness	0.419	0.426	0.405
BFI-Openness	0.460	0.435*	0.448
<i>Other Survey Variables</i>			
Ethic Score	0.742	0.689**	0.716
Integrity Score	0.446	0.432	0.434
People take advantage	0.555	0.508**	0.481***
Amount to put in risky option	0.398	0.471**	0.447*
Extreme Sports	0.074	0.082	0.080
Amount to keep in UG	0.509	0.534	0.583***
<i>Altruism Variables</i>			
Donation to charity	0.189	0.178	0.100***
Donation to researcher	0.148	0.132	0.075***
No of observation	307	235	350

Notes: Mean values are represented in the table. A two-sided *t*-test is used where Column 2 and Column 3 are compared with Column 1, separately. Stars indicate statistical significance as follows: **p* < 0.10, ***p* < 0.05 and ****p* < 0.01.

For the high incentive task, this relationship is significantly different for those who are never a detectable liar as compared to the other two groups (*p* < 0.01), which suggests consistency across tasks.

Table 4 classifies subjects based on their dishonest behavior in the two different dishonesty tasks and compares their personal and other characteristics. Subjects are classified as ‘never a liar’ if they sent an honest message in both the low and high incentive versions of the sender–receiver game, ‘once a liar’ if they sent only one dishonest message in either the low or high incentive task and ‘always a liar’ if they sent a dishonest message in both. The table compares the ‘never a liar’ group of subjects with the ‘once a liar’ and ‘always a liar’ groups separately and reports significance levels from a two-sided *t*-test. We note that people who sent more dishonest messages in the sender–receiver game report more matrices being solved in the matrix puzzle game. Subjects

who always sent an honest message in the sender–receiver game reported 4.583 (4.547) solved matrices on average in the low (high) incentive matrix task, whereas subjects who sent a dishonest message once reported 5.183 (5.306) solved matrices and those who sent a dishonest message twice reported 5.9 (5.931) solved matrices. We also note that the proportion of subjects who report more than 10 matrices solved in the low (high) matrix puzzle game (the detectable liars) is higher for the groups who sent at least one dishonest message in the sender–receiver game than for the group of individuals who always sent an honest message. Putting this all together we have our following result which confirms Hypothesis 5:

Result 5: A subject who lies in the sender–receiver game is more likely to be a detectable liar in the matrix puzzle game and vice versa.

In order to check whether Hypothesis 6 (the moral balancing argument) holds, we compare the level of lying to our two donation variables. After dishonesty tasks in both waves of the experiment, we asked subjects how much they would like to donate from their bonus to *MacMillan Cancer Support*²⁸ and also how much they could wish to donate to the team of researchers to conduct more sessions in the experiment. We opted for two different measures in an attempt to disentangle reciprocity which seems more likely to apply to donations to the researcher, from general altruism. In the first instance, we opted to be as open as possible to the possibility that moral balancing applies, biasing our results as heavily as we could in that direction: then if Hypothesis 6 fails we would have the strongest possible case against moral balancing. On that basis, we consider subjects who lied the most in the sender–receiver game (in which lying is most easily detectable) and who donate more money *either* to the charity or to the researchers (despite the fact that this could entail reciprocity rather than moral balancing) to be attempting to balance out their dishonesty with moral action. According to Table 4, while the mean level of donations made to charity is 18.9% for the group of people who did not send a dishonest message in the sender–receiver game, it is 17.8% for those who lied once and 10.0% for those who always lied. This amounts to a statistically significant decrease in the donation made to charity and to the researchers for the people who always sent a dishonest message in the sender–receiver game (in both cases, $p < 0.01$). Our results suggest that the most extreme liars, rather than attempting to morally balance out their actions, instead donate *less* to charity *and* to the researchers than people who did not lie in the sender–receiver game. Moreover, as discussed there is a strong correlation between our two dishonesty tasks (Result 5) which seems to rule out the possibility that participants adjust their actions in the second experimental task to compensate for any earlier dishonesty in the first experimental

²⁸We chose Macmillan Cancer Support because it was the charity that got the most donations when we were designing our experiment. We understand that people's willingness to donate might have been different if we had chosen a different charity.

task. Therefore, Hypothesis 6 (the moral balancing argument) is not supported by our data:²⁹

Result 6: The moral balancing argument does not hold in our sample.

We also have a number of other tests and scales that we can use in an attempt to tease out any interesting further results though these do not link to any particular hypotheses. First, we consider the results from a short version of the Big Five Inventory (Rammstedt and John, 2007). Comparing the subjects who never sent a dishonest message in the sender–receiver game with the groups who sent a dishonest message once or twice, we do not observe any prevalent pattern in terms of personality. However, there are some important differences among groups. A subject who lied once in the sender–receiver game appears to be more extravert ($p < 0.05$) and open ($p < 0.10$) than subjects who never lied. Also, subjects who always sent a dishonest message in the sender–receiver game are more conscientious ($p < 0.05$) and less neurotic ($p < 0.10$) than subjects who never sent a dishonest message.

Second, we asked subjects questions about the justifiability of unethical actions and created an ethics score which is higher if they answer that more of these actions are unjustifiable. We observe that people who lied once in the sender–receiver game have a lower ethics score than people who never lied ($p < 0.05$). However, this relationship does not hold true for subjects who always lied. We also asked subjects whether they had engaged in some unethical actions such as avoiding public transport fares within the last 12 months and created an integrity score which is higher if they had engaged in fewer of these activities. The results do not show any significant difference between groups. In our test of risk aversion, we observe that subjects who never lied in the sender–receiver game invest less money in the risky option, displaying more risk-averse behavior, than the other two groups of subjects who lied once ($p < 0.05$) and twice ($p < 0.10$). They also believe more in the idea that people are likely to be fair rather than trying to take advantage of others than the other two groups. Finally, subjects who always lied in the sender–receiver game retained more money in the ultimatum game than subjects who opted never to lie in the sender–receiver game ($p < 0.01$).

Discussion

What happens when individuals' own dishonesty becomes more salient? We asked this question in the introduction, and our results seem to give a clear but subtle answer: it depends upon the context. In a real effort task with individual incentives, saliency reduces dishonesty. This effect persists, albeit with a smaller effect size, when we purposefully introduce competitive incentives to this task in wave 2. On the other hand, in a competitive environment in which subjects could earn more by lying to their counterparts, priming them to think more about honesty pushes them toward becoming more dishonest. This might be viewed as a form of acceptance, spurred on by the psychological costs of cognitive dissonance. We might also think about the psychological

²⁹ Given the lower level of donations to the researchers, nor does there appear to be any reciprocity stemming from those who lied the most.

processes which make environments with competitive incentives so different. In our discussion on the comparison of tasks and in our analysis plan, we make the case for competitive environments changing the social norms associated with lying. However, it might be interesting to probe deeper into this question in future work on the psychology of competition.

Our findings show remarkable consistency between those who lie in one task and in the other, notwithstanding our findings on the importance of context. We also see what might be called 'partial lying' as a common outcome: the situation where participants lie but not to the full extent possible. Within the broader literature, there have been numerous attempts to explain the phenomena of 'partial cheaters' as a battle between the competing forces of cognitive dissonance (the desire to maintain a positive self-image) and moral balancing.

The literature has produced some contradictory results that surround which of these two forces is considered to be dominant.³⁰ In many ways, our findings offer an explanation for any divergences in the literature: much depends upon context.

Attempting to derive policy implications from our findings, we would urge caution when attempting to use salience as a weapon against dishonesty. This can easily backfire especially in the presence of competitive incentives, forcing dishonest individuals to accept their true nature in order to avoid cognitive dissonance and making it easier for them to behave dishonestly in the future.

Competitive environments do of course exist in a variety of different important real-world situations, not just in economics or management: politicians attempting to defeat or outmaneuver their political opponents by winning elections, celebrities trying to garner more popularity and fame than their rivals, or newspapers attempting to drive competitors out of business. In each of these cases, there are many good reasons to highlight the dishonesty of others, but we cannot assume that it will also result in a period of moral balancing by those caught out in a lie. However, our results also indicate that in the presence of individual incentives only, salience of dishonesty can lead to improved behavior. This can justify the use of topical forms of rehabilitation therapy for

³⁰To provide a partial overview, Levit (2006) argues that the existence of partial liars who do not maximize their material payoff by lying could be explained by a model in which people receive internal rewards for being honest. Mazar *et al.* (2008) and Fischbacher and Föllmi-Heusi (2013) explain this phenomenon as the result of the material benefit of cheating conflicting with the desire to avoid cognitive dissonance which drives people to behave honestly in an effort to see themselves as moral. The result of this conflict determines the action which is susceptible to one's self-control or contextual framework. Abeler *et al.* (2019) support this idea by stating that a preference for being seen as honest and a preference for being honest are the main motivations for truth-telling. Monin and Miller (2001), Sachdeva *et al.* (2009) suggest an alternative framework which focuses on 'moral balancing/moral licensing' which relates to an individual's attempts to maintain an adequate level of morality in the light of past misdeeds. In this paradigm, if a person's moral self-image drops below a personal standard (that she determines through her own beliefs about her morality and her beliefs about how others perceive her morality) she would react by engaging in moral compensation to reduce cognitive dissonance (Gneezy *et al.*, 2014). This also suggests that if her moral self-image rises above her personal standards, then she might become prone to behave immorally since they have done their good deed for the day already (Mazar and Zhong, 2010; Young *et al.*, 2012; Ploner and Regner, 2013). On the other hand, some studies have found that when people are primed to focus on a positive self-identity, it can motivate them to behave accordingly in accordance with what is often called 'moral reinforcement' (Stone and Cooper, 2001; Nelson and Norton, 2005; Young *et al.*, 2012).

those caught undertaking criminal activity, for instance, self-reflection, or interacting with victims in an attempt to prime individuals to think more about their past behavior. For example, in the UK surveys have suggested that where offenders were asked to meet with their victims, the result was a reduction in reoffending of around 14%.³¹ It can also provide some justification for journalistic revelations about bad behavior, though we note that this can easily backfire where environments are competitive.

Supplementary material. To view supplementary material for this article, please visit <https://doi.org/10.1017/bpp.2025.10009>.

Competing interests. None.

Ethical approval. Ethical approval reference ECONPGR 05/18. AEA RCT registrations AEARCTR-0005142 (wave 1) and AEARCTR-0005955 (wave 2).

References

- Abeler, J., A. Becker and A. Falk (2014), 'Truth-telling: a representative assessment', *Journal of Public Economics*, **113**: 96–104.
- Abeler, J., D. Nosenzo and C. Raymond (2019), 'Preferences for truth-telling', *Econometrica*, **87**(4): 1115–1153.
- Aveyard, M. E. (2014), 'A call to honesty: extending religious priming of moral behaviour to Middle Eastern muslims', *PLOS One*, **9**(7): e99447.
- Aydogan, G., A. Jobst, K. D'Ardenne, N. Müller and M. G. Kocher (2017), 'The detrimental effects of oxytocin-induced conformity on dishonesty in competition', *Psychological Science*, **28**(6): 751–759.
- Ba'si'c, Z. and S. Quercia (2020), 'The influence of self and social image concerns on lying', *MPI Collective Goods Discussion Paper* **2020/18**.
- Benabou, R. and J. Tirole (2016), 'Mindful economics: the production, consumption, and value of beliefs', *Journal of Economic Perspectives*, **30**(3): 141–164.
- Bender, J., A. M. O'Connor and A. D. Evans (2018), 'Mirror, mirror on the wall: increasing young children's honesty through inducing self-awareness', *Journal of Experimental Child Psychology*, **167**: 414–422.
- Buccioli, A. and M. Piovesan (2011), 'Luck or cheating? a field experiment on honesty with children', *Journal of Economic Psychology*, **32**: 73–78.
- Cohn, A., M. A. Marechal and T. Noll (2015), 'Bad boys: how criminal identity salience affects rule violation', *The Review of Economic Studies*, **82**(4): 1289–1308.
- Cohn, A., M. A. Marechal, D. Tannenbaum and C. L. Zu'nd (2019), 'Civic honesty around the globe', *Science*, **365**: 70–73.
- Dickerson, C. A., R. Thibodeau, E. Aronson and D. Miller (1992), 'Using cognitive dissonance to encourage water conservation', *Journal of Applied Social Psychology*, **93**: 841–854.
- Diener, E. and M. Wallbom (1976), 'Effects of self-awareness on antinormative behavior', *Journal of Research in Personality*, **10**: 107–111.
- Dimant, E., G. A. Van Kleef and S. Shalvi (2020), 247–266.
- Dreber, A. and M. Johannesson (2008), 'Gender differences in deception', *Economics Letters*, **99**(1): 197–199.
- Erat, S. and U. Gneezy (2012), 'White lies', *Management Science*, **58**(4): 723–733.
- Falk, A. (2021), 'Facing yourself – a note on self-image', *Journal of Economic Behavior & Organization*, **186**: 724–734.
- Faravelli, M., L. Friesen and L. Gangadharan (2015), 'Selection, tournaments, and dishonesty', *Journal of Economic Behavior & Organization*, **110**: 160–175.
- Fenigstein, A. and M. P. Levine (1984), 'Self-attention, concept activation and the causal self', *Journal of Experimental Social Psychology*, **20**(3): 231–245.

³¹See <https://www.theguardian.com/uk/2012/oct/30/offenders-meet-victims-plan>

- Fischbacher, U. and F. Föllmi-Meuli (2013), 'Lies in disguise: an experimental study on cheating', *Journal of European Economic Association*, **11**: 525–547.
- Fosgaard, T. R., L. G. Hansen and M. Piovesan (2013), 'Separating will from grace: an experiment on conformity and awareness in cheating', *Journal of Economic Behavior & Organization*, **93**: 279–284.
- Gennaioli, N. and A. Shleifer (2010), 'What comes to mind', *Quarterly Journal of Economics*, **125**(4): 1399–1433.
- Gerlach, P., K. Teodorescu and R. Hertwig (2019), 'The truth about lies: a meta-analysis on dishonest behavior', *Psychological Bulletin*, **145**(1): 1.
- Gino, F., S. Ayal and D. Ariely (2009), 'Contagion and differentiation in unethical behavior', *Psychological Science*, **20**(3): 393–398.
- Gino, F., E. L. Krupka and R. A. Weber (2013), 'License to cheat: voluntary regulation and ethical behavior', *Management Science*, **59**(10): 2187–2203.
- Gino, F., M. I. Norton and R. A. Weber (2016), 'Motivated bayesians: feeling moral while acting egoistically', *Journal of Economic Perspectives*, **30**(3): 189–212.
- Gneezy, U. (2005), 'Deception: the role of consequences', *American Economic Review*, **95**(1): 384–394.
- Gneezy, U., A. Imas and K. Madarász (2014), 'Conscience accounting: emotion dynamics and social behavior', *Management Science*, **60**(11): 2645–2658.
- Gneezy, U., A. Kajackaite and J. Sobel (2018), 'Lying aversion and the size of the lie', *American Economic Review*, **108**(2): 419–453.
- Gneezy, U. and J. Potters (1997), 'An experiment on risk taking and evaluation periods', *Quarterly Journal of Economics*, **112**(2): 631–645.
- Houser, D., S. Vetter and J. K. Winter (2012), 'Fairness and cheating', *European Economic Review*, **56**: 1645–1655.
- Jordan, J., E. Mullen and J. K. Murnighan (2011), 'Striving for the moral self: the effects of recalling past moral actions on future moral behavior', *Personality and Social Psychology Bulletin*, **37**(5): 701–713.
- Kajackaite, A. and U. Gneezy (2017), 'Incentives and cheating', *Games and Economic Behavior*, **102**: 433–444.
- Kerschbamer, R., D. Neurerer and A. Gruber (2019), 'Do altruists lie less?', *Journal of Economic Behavior & Organization*, **157**: 560–579.
- Kettle, S., M. Hernandez, M. Sanders, O. Hauser and S. Ruda (2017), 'Failure to captcha attention: null results from an honesty priming experiment in Guatemala', *Behavioral Sciences*, **7**(28): 28.
- Kocher, M. G., S. Schudy and L. Spantig (2018), 'I lie? we lie! why? Experimental evidence on a dishonesty shift in groups', *Management Science*, **64**(9): 3995–4008.
- Kristal, A., A. V. Whillans, M. H. Bazerman, F. Gino, L. L. Shu, N. Mazar and D. Ariely (2020), 'Signing at the beginning versus at the end does not decrease dishonesty', *Proceedings of the National Academy of Sciences*, **117**(13): 7103–7107.
- Levit, S. D. (2006), 'White-collar crime writ small: a case study of bagels, donuts and the honor system', *Academic Economic Review*, **96**: 290–294.
- Mazar, N., O. Amir and D. Ariely (2008), 'The dishonesty of honest people: a theory of self-concept maintenance', *Journal of Marketing Research*, **45**(6): 633–644.
- Mazar, N. and C. Zhong (2010), 'Do green products make us better people?', *Psychological Science*, **21**: 494–498.
- McCabe, D. L. and L. K. Trevino (1993), 'Academic dishonesty: honor codes and other contextual influences', *Journal of Higher Education*, **64**: 522–538.
- McCabe, D. L. and L. K. Trevino (1997), 'Individual and contextual influences on academic dishonesty: a multicampus investigation', *Research in Higher Education*, **38**: 379–396.
- Mead, N. L., R. F. Baumeister, F. Gino, M. E. Schweitzer and D. Ariely (2009), 'Too tired to tell the truth: self-control resource depletion and dishonesty', *Journal of Experimental Social Psychology*, **45**(3): 594–597.
- Merritt, A. C., D. A. Effron and B. Monin (2010), 'Moral self-licensing: when being good frees us to be bad', *Social and Personality Psychology Compass*, **4**(5): 344–357.
- Monin, B. and D. Miller (2001), 'Moral credentials and the expression of prejudice', *Journal of Personality and Social Psychology*, **81**(1): 33–43.
- Mullen, E. and B. Monin (2016), 'Consistency versus licensing effects of past moral behavior', *Annual Review of Psychology*, **67**: 363–385.
- Nelson, L. and M. Norton (2005), 'From student to superhero: situational primes shape future helping', *Journal of Experimental Social Psychology*, **41**: 423–430.

- Ploner, M. and T. Regner (2013), 'Self-image and moral balancing: an experimental analysis', *Journal of Economic Behavior & Organization*, **93**: 374–383.
- Pruckner, G. J. and R. Sausgruber (2013), 'HONESTY ON THE STREETS: a FIELD STUDY ON NEWSPAPER PURCHASING', *Journal of the European Economic Association*, **11**: 661–679.
- Rabin, M. (1994), 'Cognitive dissonance and social change', *Journal of Economic Behaviour and Organization*, **23**: 177–194.
- Rammstedt, B. and O. P. John (2007), 'Measuring personality in one minute or less: a 10item short version of the big five inventory in English and German', *Journal of Research in Personality*, **41**: 203–212.
- Rosenbaum, S. M., S. Billinger and N. Stieglitz (2014), 'Let's be honest: a review of experimental evidence of honesty and truth-telling', *Journal of Economic Psychology*, **45**: 181–196.
- Sachdeva, S., R. Iliev and D. Medin (2009), 'Sinning saints and saintly sinners: the paradox of moral self-regulation', *Psychological Science*, **20**(4): 523–528.
- Shu, L. L., F. Gino and M. H. Bazerman (2011a), 330–349.
- Shu, L. L., F. Gino and M. H. Bazerman (2011b), 'Dishonest deed, clear conscience: when cheating leads to moral disengagement and motivated forgetting', *Personality and Social Psychology Bulletin*, **37**(3): 330–349.
- Shu, L., N. Mazar, F. Gino and M. Bazerman (2011), 'When to sign on the dotted line? signing first makes ethics salient and decreases dishonest self-reports', *Harvard Business School: Negotiation, Organizations Markets Unit Working Paper Series*, **45**(6): 633–644.
- Stone, J. and J. Cooper (2001), 'A self-standards model of cognitive dissonance', *Journal of Experimental Social Psychology*, **37**: 228–243.
- Stoop, J. (2014), 'From the lab to the field: envelopes, dictators and manners', *Experimental Economics*, **17**: 304–313.
- Sutter, M. (2009), 'Deception through telling the truth?! Experimental evidence from individuals and teams', *The Economic Journal*, **119**: 47–60.
- Valdesolo, P. and D. DeSteno (2007), 'Moral hypocrisy: social groups and the flexibility of virtue', *Psychological Science*, **18**(8): 689–690.
- Valdesolo, P. and D. DeSteno (2008), 'The duality of virtue: deconstructing the moral hypocrite', *Journal of Experimental Social Psychology*, **44**: 1334–1338.
- Verschuere, B., E. H. Meijer, A. Jim, K. Hoogesteyn, R. Orthey, R. J. McCarthy, J. J. Skowronski, O. A. Acar, B. Aczel, B. E. Bakos et al. (2018), 'Registered replication report on Mazar, Amir, and Ariely (2008)', *Advances in Methods and Practices in Psychological Science*, **1**(3): 299–317.
- Whiteley, P. (2012), 'Are people in the UK becoming more dishonest', <https://www.bbc.co.uk/news/av/uk-16714502> [17 June 2025].
- Young, L., A. Chakroff and J. Tom (2012), 'Doing good leads to more good: the reinforcing power of a moral self-concept', *Review of Philosophy and Psychology*, **3**(3): 325–334.