Cognitive Flexibility or Moral Commitment? Evidence of Anticipated Belief Distortion*

Silvia Saccardo† and Marta Serra-Garcia‡

December 2020

Abstract

Do people anticipate the conditions that enable them to manipulate their beliefs when confronted with unpleasant information? We investigate whether individuals seek out the "cognitive flexibility" needed to distort beliefs in self-serving ways, or instead attempt to constrain it, committing to unbiased judgment. Experiments with 6500 participants, including financial and legal professionals, show that preferences are heterogeneous: over 40% of advisors prefer flexibility, even if costly. Actively seeking flexibility does not preclude belief distortion. Individuals anticipate the effects of cognitive flexibility and their choice to pursue it responds to incentives, suggesting some sophistication about the cognitive constraints to belief distortion.

JEL Classification: D83, D91, C91.

Keywords: Belief distortion, morality, sophistication, commitment, experiments.

*We would like to thank Johannes Abeler, Saurabh Bhargava, Christine Exley, Laura Gee, David Huffman, Alex Imas, Michel Marechal, Kirby Nielsen, Ricardo Perez-Truglia, Peter Schwardmann, Jeroen van de Ven, Joel van der Weele, Lise Vesterlund, Roberto Weber, Florian Zimmerman, and participants at several conferences and workshops for helpful comments and suggestions. We would also like to thank Ben Schenk for excellent research assistance.

†Department of Social and Decision Sciences, Carnegie Mellon University and CESifo. Email: ssaccard@andrew.cmu.edu

‡Rady School of Management, UC San Diego and CESifo. Email: mserragarcia@ucsd.edu.

1 Introduction

The fundamental desire to preserve a positive identity often leads individuals to engage in motivated reasoning as a way to protect valued beliefs (e.g., Kunda, 1990; Bénabou and Tirole, 2006, 2011, 2016; Köszegi, 2006). Motivated beliefs can explain phenomena such as managerial overconfidence (e.g., Malmendier and Tate, 2005, 2008), partisan polarization (Kahan, 2013), or collective denial of wrongdoing in organizations (e.g., Bénabou, 2013). To protect their self-view, individuals often avoid inconvenient information (e.g., Dana, Weber and Kuang, 2007; Golman et al., 2017). However, sustaining motivated beliefs is particularly challenging when individuals receive feedback that threatens these beliefs. To that end, individuals engage in ex-post signal distortion using a variety of technologies: They underweight informative signals (e.g., Eil and Rao, 2011; Sharot et al., 2011; Möbius et al., 2013), interpret them self-servingly (e.g., Kunda, 1990; Di Tella et al., 2015) or selectively forget them over time (e.g., Zimmerman, 2020; Huffman et al., 2020). However, there are cognitive limits to people's ability to distort informative signals: Belief distortion can be enabled or constrained by contextual cues (Epley and Gilovich, 2016; Sloman, Fernbach, and Hagmayer, 2010). An important open question about motivated cognition is whether individuals anticipate the conditions under which they can more easily manipulate their beliefs when confronted with unpleasant information. And if so, would they attempt to constrain anticipated self-deception, or would they rather seek out ways to more easily dismiss a *potentially inconvenient truth?*

We investigate this question in the domain of moral behavior (see, for example, Abeler et al., 2019; Cohn et al., 2019). Individuals often give in to the temptation to profit from self-serving behavior but manipulate their beliefs to protect their self-view as moral.[1] When directly confronted with undesirable information, altering beliefs requires "cognitive flexibility": the cognitive ability to dismiss, underweight, or flexibly interpret informative signals. If given the choice, individuals may antic-

[1] A large literature suggests that self-serving behavior is more likely when decisions can be rationalized by exploiting ambiguity or subjectivity in the decision environment (e.g., Hsee, 1996; Konow, 2000; Haisley and Weber, 2010; Exley, 2015; Di Tella et al., 2015; Shalvi et al., 2011; Shalvi et al., 2015; Gneezy, Saccardo, and van Veldhuizen, 2018; Gneezy et al., 2020; Falk, Neuber, and Szech, 2020), by avoiding information about how their choices affect others (e.g., Dana, Weber and Kuang, 2007; Larson and Capra, 2009; Grossman, 2014; Grossman and van der Weele, 2017; Serra-Garcia and Szech, 2019), or by conveniently forgetting unpleasant news (Kouchaki and Gino, 2016; Saucet and Villeval, 2019; Carlson et al., 2020). These rationalizations are sought out by agents with self-image concerns, who desire to preserve a positive self-view (see, for example, Quattrone and Tversky, 1984; Bodner and Prelec, 2003; Bénabou and Tirole, 2006; Mijovic-Prelec and Prelec, 2010).

ipate the conditions that provide cognitive flexibility, and exploit them to preserve a positive self-view in spite of a moral transgression. Or they may actively choose to constrain cognitive flexibility, committing to accurate beliefs as a way to uphold their morals and minimize the temptation to transgress.

We study preferences for cognitive flexibility or moral commitment in a series of experiments where participants ($N = 6,526$) face a potential moral dilemma and can choose the order with which they receive a sequence of signals. In many moral dilemmas, individuals receive information about what is in their best interest as well as information about what is best for another party. The order with which this information is presented can constrain cognitive flexibility: Making a first assessment of what is best for another party without knowing one's own incentives might commit individuals to a first unbiased judgment (e.g., Goldin and Rouse, 2000), restricting the ability to justify self-serving behavior once new information is received (e.g., Babcock et al., 1995; Gneezy et al., 2020; Schwardmann et al., 2020). One important example is fiduciary relationships, where experts —financial or investment advisors, attorneys, accountants, or corporate board members—have the ethical responsibility to put their clients' interests above their own, but may manipulate their beliefs to justify actions that violate this duty. When new investment funds, insurance policies, or financial products are issued, financial advisors may actively commit to accurate beliefs by first assessing whether the new products are appropriate for their client, or may rather seek out the cognitive flexibility needed to distort their beliefs by first looking up which product yields them the highest commission.

In our experiments, an advisor recommends one of two products to an uninformed client and faces a potential conflict of interest. The payoff distribution of one of the products, which we refer to as "quality," is uncertain. The advisor receives two pieces of information: a signal about the quality of the uncertain product and information about her private incentive (i.e., which product the advisor is incentivized to recommend). All advisors receive both pieces of information but can choose the order with which they receive information. Seeing the incentive first may increase the salience of this piece of information, drawing attention to it and helping advisors underweight subsequent information that is in conflict with their desired recommendation.[2] Conversely, assessing quality first might increase the salience of

[2]The important role of attention and salience in economic choices has been shown in Gabaix et al. (2006), Chetty, Looney and Kroft (2009), Bordalo, Gennaioli, and Shleifer, (2012, 2013), Koszegi and Szeidl (2013) and Schwartzstein (2014), among others. There is also work on motivated attention (e.g., Ditto and Lopez, 1992; Tasoff and Madarasz, 2009; Fehr and Rangel, 2011; Sicherman et al., 2016; Golman et al., 2019).

this signal, reducing the cognitive flexibility needed for engaging in "reality denial" in case of a conflict of interest. According to Bénabou and Tirole (2016), reality denial—the failure to update beliefs properly in response to bad news—is one of the main strategies used to protect valuable beliefs.

We begin by demonstrating that, in line with this hypothesis and prior work, when the sequence of signals is exogenously assigned, advisors are more likely to make recommendations that are in the client's best interest when the signal about quality is assessed first. When identical information is presented in the alternative order, with information about incentives coming first, recommendations are much more biased toward the incentive. This is not a purely cognitive bias: advice is not affected by the order of information when advisors' interests are aligned with those of the client.

Our main experiment investigates preferences, recommendations, and beliefs when advisors have the option to *choose* the sequence of information. Do advisors seek out cognitive flexibility, preferring to know about their incentives before seeing the signal about quality? Or do they prefer to see the signal first as a commitment to providing unbiased advice to the client? We use data from a sample of professionals employed in the finance (including insurance) and legal services industries, who are typically more exposed to conflicts of interest, and from a general (convenient) sample of online participants. Across both samples, we find substantial heterogeneity in preferences, which are split between cognitive flexibility and commitment. If the choice is costless, 42% of advisors in the convenience sample and 52% of advisors in the sample of professionals commit to more accurate beliefs (with the remaining 58% and 48%, respectively, seeking out cognitive flexibility). Preferences for cognitive flexibility are positively correlated with selfishness, consistent with the idea that selfish advisors are the ones who may seek out the cognitive flexibility needed to recommend the product that is in their best interest without damaging their self-image. Notably, a substantial fraction of participants (40%) prefers flexibility even when there is a financial incentive to choose commitment, which provides suggestive evidence for the hypothesis that advisors anticipate the benefits of belief distortion.

Overall, preferences for flexibility or commitment result in different recommendation patterns, with advisors who choose to see their incentives first being overall more likely to recommend the incentivized product than those who commit to seeing the signal about quality first. A prominent hypothesis in the philosophical discourse on self-deception suggests that actively *pursuing* flexibility prevents individuals from

subsequently being successful at using this flexibility to self-deceive.[3] Our findings suggest that this is not the case. Using a random assignment mechanism we compare, conditional on preferences, advisors' recommendations when they receive information in their preferred order and when they do not. We find that, conditional on seeking out cognitive flexibility, advisors who actually get it—receiving information about their incentive first—are more likely to behave self-servingly. This finding suggests that actively seeking out cognitive flexibility does not undermine people's ability to distort their beliefs and rationalize self-serving behavior. When advisors do not face a conflict of interest, the effect of information order is substantially smaller, suggesting that belief distortion is motivated. In support of this evidence, data on beliefs show that advisors who demand cognitive flexibility are more likely to engage in reality denial, failing to update their beliefs after observing a signal about quality that is in conflict with their incentives.

The finding that advisors are willing to incur a cost to enjoy the benefit of cognitive flexibility is suggestive that advisors anticipate the effect of information order on belief updating. However, preferences for information in this experiment may be driven by factors other than the desire to amplify or mitigate self-serving judgment. We provide further suggestive evidence of anticipation in two additional experiments. First, we leverage a forecasting experiment (Della Vigna and Pope, 2018; Della Vigna, Pope, and Vivalt, 2019) to test whether a different group of individuals anticipates the effect of different information sequences on behavior. We find that a large fraction of participants predicts that getting to see the incentive first increases recommendations of the incentivized product. The average predicted effect is not significantly different from the actual effect we observe in our experiments. This finding suggests that many individuals anticipate the effects of cognitive flexibility or commitment on behavior.

Second, we test whether preferences for cognitive flexibility or commitment respond to incentives. Most models of motivated cognition assume that belief distortion is driven by incentives (e.g., Bénabou and Tirole, 2011; Brunnermeier and Parker, 2005). If individuals anticipate the conditions that facilitate belief distortion, we would expect them to be responsive to the potential gains from manipulating their beliefs in response to signals of quality that are in conflict with their incentives. If, instead, preferences for flexibility or commitment are driven by other

[3]The question of whether individuals can intend to distort their beliefs and effectively self-deceive without rendering their intention ineffective has been widely debated in the philosophy literature (Mele, 1987, and 2001; Bermúdez, 2000; see also Mijovic-Prelec and Prelec, 2010).

factors, such as a mere desire to fulfill curiosity, the potential gains from belief distortion should be irrelevant. Our data show that, when we reduce the potential gains from distorting beliefs, thereby reducing advisors' incentives to demand cognitive flexibility, very few advisors (13%) demand to see their incentives first. Taken together, our experiments suggest that individuals anticipate what type of information sequence constrains or enables the cognitive flexibility necessary for engaging in belief distortion, and there is substantial heterogeneity in such preferences.

Our research contributes to a growing literature on the malleability of moral behavior (Konow, 2000; Haisley and Weber, 2010; Moore, Tanlu, and Bazerman, 2010; Trivers, 2011; Bénabou, 2015; Exley, 2015; Bénabou, Falk and Tirole, 2018; Gino, Norton and Weber, 2016; Epley and Gilovich, 2016; Exley and Kessler, 2019) and on the channels through which individuals form and maintain motivated beliefs. One channel is information avoidance: Individuals strategically avoid information that could bear negative news, as first shown by Dana et al., 2007 (see also Oster, Shoulson, and Dorset, 2013; Grossman and van der Weele, 2017; Ganguly and Tasoff, 2017; Golman et al., 2019; for a review see Golman, Haggman and Loewenstein, 2017). However, as highlighted by Bénabou (2015), in many situations unwelcome information cannot be avoided—for example, when employees receive performance evaluations or, more similar to our experiments, when fiduciaries receive information about the best course of action for their client. Research has shown that, when "bad" news cannot be escaped, individuals underweight or distort aversive signals, and forget them over time. These behaviors are easier if the environment provides scope for such rationalizations.

Our paper is the first to show that when avoiding information is not feasible, individuals anticipate that subtle changes to the way information is received can enable or mitigate reality denial. We show that a substantial fraction of individuals are willing to pay for cognitive flexibility, but only when the gains from belief distortion are large enough. Notably, our findings illustrate that *actively pursuing* cognitive flexibility does not limit the extent of self-deception via reality denial. These results provide new empirical evidence regarding one of the key puzzles in the philosophical discourse on self-deception: whether individuals can intend to deceive themselves without rendering their intentions ineffective (Mele, 1987 and 2001). Or, in other words, whether individuals can be somewhat conscious of belief distortion and still effective at manipulating their beliefs. Taken together, our findings provide the first evidence of sophistication in this form of self-deception.

This work has important implications for the broader literature on information asymmetries and conflicts of interest (e.g., Darby and Karni, 1973), whereby fiduciaries can exploit their private information to serve their own interests (Crawford and Sobel, 1982; Pitchik and Schotter, 1987; Bénabou, 2013; Sobel, 2020). It also speaks to work on the behavioral determinants of corruption (e.g., Weisel and Shalvi, 2015; Malmendier and Schmidt, 2017). Our findings suggest that professionals and laypeople are aware of some of the conditions that facilitate ex-post signal distortion, with a considerable fraction of advisors choosing to exploit them. Conflicts of interest permeate both corporations and governments. If those who face potential conflicts of interest play a role in designing rules and regulations, many institutions may be designed in a way that maximizes personal enrichment while protecting self-image. Fiduciaries, such as corporate boards or public officials, often take an active part in the design of the institutional arrangements that govern their own behavior. These arrangements might thus be designed ex-ante to consider all relevant information needed for optimal decisions (e.g, company financial performance measures), but to present information in a way that provides maximum flexibility to pursue private interests while preserving the belief that they are ethical.

2 Experimental Design

Our aim is to test whether individuals attempt to seek out the cognitive flexibility needed for behaving self-servingly without harming their self-image, or rather attempt to constrain cognitive flexibility in order to uphold their morals, a form of commitment to accurate beliefs and moral behavior. Further, we are also interested in testing how the *choice* to pursue or restrain cognitive flexibility affects behavior and beliefs in the face of unwelcome information. Studying this question requires an environment where (i) individuals are tempted to put their own interests above those of another party, and (ii) that provides them with the cognitive flexibility needed to pursue private gains while maintaining a self-image as moral. Further, it requires an environment where (iii) individuals can actively pursue cognitive flexibility (or, conversely, mitigate it) and (iv) that allows studying the effect of this active choice on subsequent behavior and beliefs. Our experiment is designed to accommodate these four features.

We study an advice game where an advisor provides product recommendations to a client and receives a signal that helps her infer what the best recommendation is.

The advisor also receives a commission, which depends on her advice. The presence of a commission leads to cases in which there is a conflict of interest such that the advisor faces a trade-off between maximizing her financial gains and providing advice that is in the best interest of the client. To earn the commission without damage to her moral self-image, the advisor may discount informative signals that are not aligned with her desired recommendation (e.g., Eil and Rao, 2011; Möbius et al., 2013), a form of reality denial (Bénabou, 2015).

Prior work has shown that the order in which decision-makers see their own incentives and evaluate the fairness or ethicality of a choice changes their scope for justifying self-interested behavior (e.g., Babcock et al., 1995; Gneezy et al., 2020).[4] Based on this work, we expect that the order in which individuals receive a sequence of two signals would affect their scope for engaging in reality denial. Seeing the incentives first could make make this information more salient, giving advisors the cognitive flexibility needed to discount subsequent signals of quality that are not in line with their incentives. Conversely, seeing the quality signal first might constrains this flexibility, as individuals update their beliefs about quality before knowing what their incentives are, making an unbiased assessment in their mind. Varying the sequence of information in our setting, as shown in Figure 1, allows us to address the aforementioned conditions (i) and (ii).

By asking advisors to choose their preferred information order—seeing the incentives first or only after having seen the signal of quality—we can study whether advisors intentionally self-select into environments that provide the cognitive flexibility to underweight undesired informative signals or instead choose to commit to an unbiased judgement, addressing condition (iii). To study the causal effect of cognitive flexibility on behavior and separate it from self-selection, we assign participants to receive their desired information order in 75% of the cases, which takes care of condition (iv). Hence, this setting provides an environment that allows us to capture all four features described above.

2.1 The Advice Game

The advisor recommends one of two products, Product A and B, to an uninformed client. Each product is presented as an urn containing five balls, as displayed in Figure 1. Product A has three \$2 balls and two \$0 balls. That is, Product A pays

[4]Note that "recency" effects by which signals seen more recently have a stronger impact on beliefs than earlier signals (see Benjamin, 2019) would predict the opposite effect.

$2 with prob 0.6, and $0 otherwise (an expected return of $1.20). Product B's payoff depends on the state, which we refer to as Product's B quality and that can be high (H) or low (L). We denote the state by $s \in \{H, L\}$, and the probability that $s = H$ is 0.5. If $s = H$, then B has four $2 balls and one $0 ball. It thus yields a higher probability of receiving $2 than Product A, as it pays $2 with prob 0.8, and $0 otherwise, for an expected return of $1.60. If $s = L$, then B has two $2 balls and three $0 balls. It thus yields a lower probability of receiving $2 than Product A, as it pays $2 with prob 0.4, and $0 otherwise, for an expected return of $0.80. The quality of Product B (s) is unknown to the advisor.

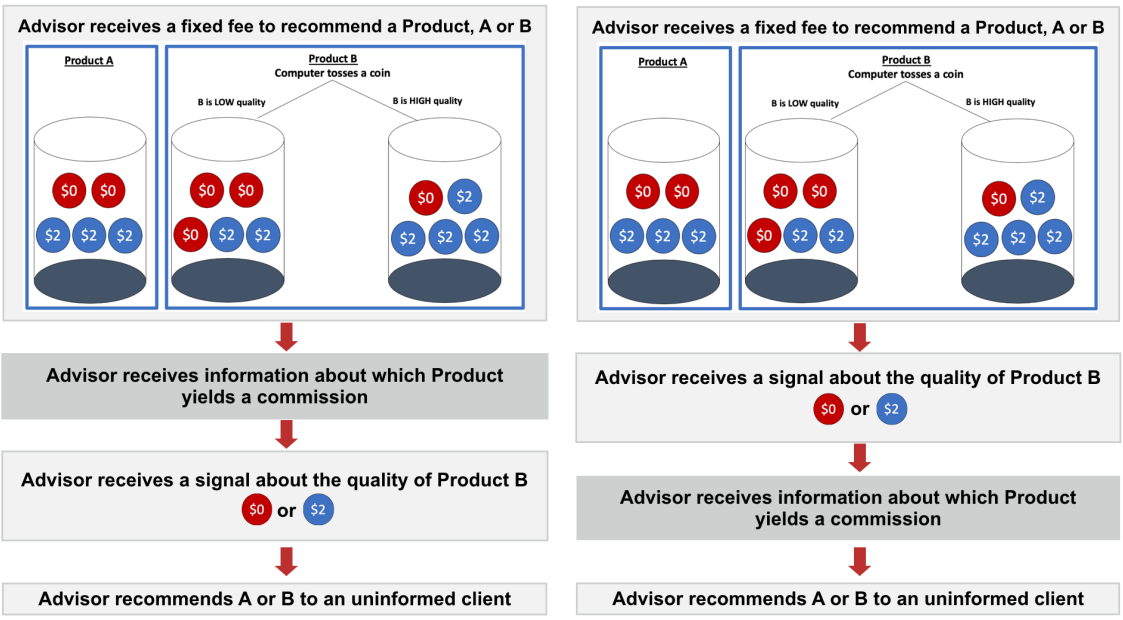
Before making the recommendation, the advisor receives a signal about the state, $\sigma \in \{H, L\}$. The signal is a ball that is randomly drawn from Product B (with replacement), which allows the advisor to update her beliefs about whether $s = H$ or $s = L$. Upon learning the signal, the advisor chooses which product to recommend to the client, Product A or Product B. After receiving the recommendation, the client chooses whether to follow the advice and is paid according to one of the balls randomly selected from the product he/she selects.

The advisor receives an incentive (commission c = $0.15), for recommending either Product A or Product B. Depending on what product is incentivized and on which signal is drawn from Product B, the advisor may face a conflict of interest. If the commission is for Product B and the signal is a $0 ball, the advisor faces a conflict between pursuing the commission (i.e., recommending Product B) and making the recommendation that is in the clients' best interest (i.e., recommending Product A). Similarly, if the commission is for Product A and the signal is a $2 ball, the advisor has to choose between maximizing her earning (i.e., recommending Product A) or making the recommendation that is best for the client (i.e., recommending Product B). In the remaining cases, the advisor does not face a conflict of interest.

2.2 The Experiments

We conduct four online experiments, summarized in Table 1.

A. The NoChoice Experiment. The goal of the first experiment is to establish that cognitive flexibility varies with the order of information. That is, in the context of our game, varying the information sequence affects the ease with which advisors recommend the product that yields them a commission when they receive a conflicting quality signal. This experiment has two treatments. In the See Incentive First



(a) See Incentive First (b) Assess Quality First

Figure 1: The Advice Game

treatment, the advisor first receives information about which product recommendation is incentivized (Figure 1a). In the Assess Quality First treatment, the advisor first sees the quality signal about B and only then learns about her incentive (Figure 1b).

In both treatments, the evaluation of the signals only occurs in the advisors' mind. However, advisors' ability to rationalize (to themselves) recommendations that yield a commission but are not in the best interest of clients may vary with the order in which information is received. In the Assess Quality First treatment, the advisor first sees the signal about quality, which allows her to learn which product has a higher expected payoff, and then sees which product would give her a commission. With this information sequence, choosing the incentivized product over the product with the higher expected payoff may damage the advisor's self-image as ethical. In the See Incentive First treatment, by contrast, after the advisor sees the incentive, she may process the signal in a self-serving way and, if the signal is in conflict with her self-interest, fail to update her beliefs about the quality of Product B at a greater extent (reality denial). This information sequence might then give advisors the cognitive flexibility needed to recommend the incentivized product with little harm to their self-view.

To isolate whether the increase in recommendations when the advisor sees the

incentive first is due to the advisors' incentive (a motivated bias), as opposed to being driven by a cognitive mechanism such as interpreting the commission as a signal of quality, demand effects, or order effects, we compare cases with and without conflict of interest.[5] If the See Incentive First treatment generates effects on advisors due to reasons that are unrelated to the conflict of interest, these effects should also appear when the signal about the quality of Product B and the commission are aligned, that is, when there is no conflict of interest.

Table 1: Experimental Design Outline

Experiment	**Treatment**	**What do advisors see first?**	**N**
Documenting Cognitive Flexibility: Information Order Affects Recommendations			
No Choice	See Incentives First	Incentive	153
	Assess Quality First	Quality Signal	148
Preferences for Information Order: Cognitive Flexibility or Moral Commitment?			
Choice	Incentive First Free	Advisor's Choice	2377
	Incentive First Free—Professionals	Advisor's Choice	713
	Incentive First Costly	Advisor's Choice	1358
Mechanisms			
Predictions	About Incentive First Costly	-	288
ChoiceStakes	Low Incentive	Advisor's Choice	486
	Intermediate Incentive	Advisor's Choice	515
	High Incentive	Advisor's Choice	488

B. The Choice Experiment. In this experiment we test whether individuals self-select into seeing the incentive first, which can provide more scope for rationalizing self-serving recommendations (i.e., preferring cognitive flexibility to altering beliefs) as opposed to self-selecting into assessing quality first (i.e., preferring commitment to more accurate beliefs). We also test how these preferences affect recommendations: do advisors who willfully pursue cognitive flexibility still have scope to distort their advice without harming their self-image?

The advisor first *chooses* whether to see the incentive first or assess quality first. The order of information is not randomly assigned but is instead based on the advisor's preference. To address the endogeneity issue that arises from self-selection, and to be able to estimate how receiving information in a given order affects recommendations, we randomize whether advisors' choices are implemented. That is,

[5]These alternative accounts have not been addressed by prior work such as Babcock et al. (1995) or Gneezy et al. (2020).

when selecting whether to see the commission or the signal of quality first, advisors are informed that their preference is implemented with 75% probability. In the remaining 25% of the cases, after making the choice, advisors are assigned to learn in the order they do not prefer.

If the advisor chooses to see the incentive first and is assigned to seeing the incentive first, she then first learns about the incentive and then sees the signal about the quality of Product B, as in Figure 1a. If the advisor chooses to assess quality first, and is assigned to seeing quality first, she first sees the signal about the quality of Product B, and thereafter learns about her incentive, as in Figure 1b.

In the *Incentive First Free* treatment, seeing the incentive first is free. We conducted this experimental treatment with a sample of individuals who work in industries in which advice is pervasive—finance (including insurance), and legal services (*Incentive First Free—Professionals*) as well as with individuals from our convenience sample (Amazon Mechanical Turk or AMT). Varying the sample allows us to compare the preferences and recommendations of individuals who are likely to deal with conflicts of interest in their professional lives to those of participants who may have such experiences less often.

When the choice is free, individuals who are indifferent may choose either information sequence. To examine whether individuals have strict preferences to see the incentive first, we also vary whether the seeing the incentive first is costly (*Incentive First Costly*). Individuals receive the equivalent of a third of their commission if they choose to see the signal of quality first. This treatment allows us to investigate whether individuals are willing to forgo incentives to see the incentive first, and possibly pursue cognitive flexibility. We conducted this treatment with the convenience sample.

This experiment allows us to answer two research questions. First, do advisors demand to see incentives first, even if costly—actively pursuing cognitive flexibility—or do they choose to see the quality signal first—committing to more accurate beliefs, and possibly constraining their ability to justify self-serving advice? Second, does actively choosing a given information sequence, especially one that can enable self-serving behavior, affect advisors' scope for biasing their advice and distorting their beliefs?

D. The Prediction Experiment. In order to be able to interpret the choices in the Choice experiment as indicative that advisors anticipate that a given sequence of information will affect the extent of cognitive flexibility, we test whether a new

set of participants can actually forecast how advisors' recommendations varied with the information sequence they were presented. In this experiment, forecasters read a summary description of the recommendation decision advisors made in the Incentive First Costly treatment of the Choice experiment and, if desired, could access the exact instructions advisors saw. In the experiment, we ask forecasters to consider the recommendation decisions of advisors who choose to see their incentives first. Across participants, we vary whether forecasters are told that these advisors were then assigned to see the incentives first or assess quality first. Forecasters are then asked to estimate the fraction of recommendations of the incentivized product. To aid participants in making their predictions, and following the approach of DellaVigna and Pope (2018), participants receive information about the counterfactual—the fraction of recommendations of the incentivized product for cases in which advisors were assigned to receive information in the opposite order. Then, we first ask forecasters to predict the direction of the effect (more, equal or fewer recommendations of the incentivized product), and then to provide their estimated fraction of recommendations. If participants anticipate that seeing the incentive first gives advisors more flexibility to provide self-serving recommendations, then we would expect to see a positive and significant gap between the two information sequences, with participants predicting a higher fraction of recommendations of the incentivized product when advisors see their incentive first.

E. The ChoiceStakes Experiment. We hypothesize that advisors anticipate that cognitive flexibility is greater when individuals see their incentive first and that their choice of seeing the incentive first is thereby driven by their desire to rationalize self-serving recommendations in the face of a potential conflict of interest. This hypothesis implies that if the gains from flexibility decrease (via a substantial decrease in the advisor's incentive), the preference to see the incentive first would drop. If instead this preference is driven by alternative mechanisms, such as curiosity, preferences for seeing the incentive first would not decrease when advisors' incentives are lower. We test this prediction in the ChoiceStakes experiment.

We vary the size of the incentive (commission) for the advisor to be either low, \$0.01 in the Low Incentive treatment, the same as in the Choice experiment, \$0.15 in the Intermediate Incentive treatment, or doubled to \$0.30 in the High Incentive treatment. Throughout, choosing to see the incentive first is costly as in the Incentive First Costly treatment. Advisors' preference is only implemented with 75% probability, following the design of the Choice experiment.

2.3 Sample and Incentives

We conducted all experiments except the Incentive First Free—Professionals treatment, on Amazon Mechanical Turk (AMT), an online platform that allows to recruit workers to complete Human Intelligence Tasks (HITs). This platform has the advantage that sample sizes can be larger than the sample sizes of typical laboratory experiments with students, allowing us to recruit a large sample of participants for the Choice experiment and further explore the mechanisms behind the choices advisors make in additional experiments. Existing research shows that classic behavioral experiments have been successfully replicated on this platform (Paolacci, Chandler, and Ipeirotis, 2010; Amir, Rand, and Gal, 2012), which is more and more commonly used by economists (e.g., DellaVigna and Pope, 2018). To be eligible for our study, workers had to have a United States IP address and had to have previously completed at least 100 tasks on AMT with a 95% approval rating. All experiments on AMT were pre-registered (see Online Appendix C for the details of each pre-registration).[6]

The sample of professionals was drawn from individuals who work in two industries in which advice is very frequent: finance and insurance, and legal services. We used Prolific Academic (Palan and Schitter, 2018) and CloudResearch (Litman et al., 2016) to target the experiment to professionals in these industries. Prolific has their own sample of participants, and we recruited as many professionals as possible within the UK, the US, and Canada. CloudResearch draws professionals from AMT, and again we recruited as many professionals based in the US as possible. We pool these two samples since choices regarding the preferred sequence of information did not vary significantly across them (p=0.3710), and recommendations did not differ either (p=0.890).

We recruited a total of 6,526 participants. There were 301 participants in the NoChoice experiment. We recruited a substantially higher number of participants in the Choice experiment (4,448 in total) in order to have sufficient statistical power to study the effect of random assignment and preference on recommendations and beliefs. There were a total of 713 in Incentive First Free—Professionals, 2,377 participants in Incentives First Free, and 1,358 in Incentive First Costly. There were 288 participants in the Prediction experiment, and 1489 in ChoiceStakes. Across all experiments, the proportion of females varied between 48% and 57% (53% in the

[6]NoChoice: aspredicted #22709; Choice: aspredicted #23272; and #42246; ChoiceStakes: aspredicted #27982; Predictions: aspredicted #37081.

sample of Professionals), and the average age between 35 and 38 (37 in the sample of Professionals). These characteristics were balanced across experiments (detailed balance checks are in Online Appendix A).[7]

Participants on AMT received a $0.50 payment for making a recommendation (and participating in a 5-7 minute study). Most participants received a $0.15 commission for recommending either Product A or Product B. In the ChoiceStakes experiment the commission were $0.01, $0.15 or $0.30, in the Low, Intermediate, and High Incentive treatments, respectively. Advisors were informed that one out of 10 advisors would be matched with a client (another AMT participant), and their advice was delivered to the clients. The Prediction experiment participants were paid $1. Participants were informed that if their predictions lay within 5 percentage points of the true value, they would receive an additional $2 payment.

To maximize recruitment, each professional was paid $1 for participating in the experiment. They were informed that one out of each 100 professionals would be randomly selected. Each selected advisor was matched to one client, and their advice delivered. Each advisor, if selected, received an incentive of $15 when recommending the incentivized product. The payoffs of Product A or Product B were scaled up to $0 or $20, to account for the one-to-one matching with the client, conditional on being selected. The aim was to make incentives more salient to professional participants, who had potentially higher opportunity costs of time. We tested whether these probabilistic incentives affected behavior significantly on AMT, and we found no evidence that they do.[8]

[7]In all experiments, we included attention checks. As pre-registered, we excluded participants who failed to pass such attention checks and participants who provided inconsistent choices in the selfishness elicitation procedure. In the Prediction Experiment, we asked participants how they decided their estimate and, as pre-registered, we excluded inattentive participants who gave answers unrelated to the study as coded by an independent rater. Details about recruitment and screening for attentive participants are provided in Online Appendix A.

[8]To test the potential effect of probabilistic incentives, we conducted the Incentive First Free treatment in two waves. We recruited a first wave with 1,324 participants, and a second wave with 1,053 participants. In the second wave, we randomized whether incentives were probabilistic and whether the incentivized product was presented on the left side or the right side of the screen. We found no effect on the preference to see the incentive first ($p > 0.1$ in both cases) or on recommendations ($p > 0.1$ in both cases), hence we pool the data and control for these design variations in all regression analyses. As pre-registered, since advisors' preferences and recommendations did not differ across waves ($p > 0.1$ for both), we pool the data in the analyses.

2.4 Procedures

All experiments were conducted online, using Qualtrics surveys. The instructions are shown in Online Appendix B. In the Choice and ChoiceStakes experiments, advisors are given two options to choose from: *"I want to learn which product recommendation gives me a \$0.15 commission (Product A or B)* ***before*** *I obtain information that helps me infer the quality of Product B"* or *"I want to learn which product recommendation gives me a \$0.15 commission (Product A or B)* ***after*** *I obtain information that helps me infer the quality of Product B."* In the Incentives First Costly treatment, and in all the treatments of the ChoiceStakes experiment, making this choice was costly, as advisors could receive an additional \$0.05 from choosing to see the incentive after seeing the signal about quality. As explained earlier, advisors knew that there was a 75% chance that their preference would be implemented. After making the choice, advisors learned whether their choice was implemented, and then proceeded to see either the commission followed by the signal, or the signal followed by the commission, with the order depending on whether their choice was implemented. Finally, advisors were prompted to make their recommendation to the client. Clients were recruited later on the same platform as that of the advisor and received no information other than the advisor's recommendation.

Our main interest is in the cases where advisors faced a conflict of interest, that is, the cases in which the signal about the quality of Product B revealed that the recommendation of the better product was *not* the recommendation that yielded a commission. Given that the signals were selected from Product B at random, we pre-determined which product yielded a commission in a way that maximized the number of cases in which advisors faced a conflict of interest. In particular, all advisors randomly assigned to having a low-quality Product B (i.e., two blue (\$2) balls and three red (\$0) balls) received a commission for recommending Product B, whose expected value was lower than that of Product A. When receiving the signal, these advisors had a 3 in 5 chance of receiving "bad news," a signal suggesting that Product B had low quality (i.e., a red ball), which created a conflict of interest. In the remaining cases, advisors received "good news," a signal that was aligned with their incentive to recommend Product B. Similarly, all advisors randomly assigned to having a high-quality Product B (i.e., four blue (\$2) balls and one red (\$0) ball) received a commission for recommending Product A, whose expected value was lower than that of Product B. By this design, 70% of advisors faced a conflict between maximizing their gains and providing advice that was in the best interest of the

client, whereas 30% of advisors did not face such a conflict.

At the end of the experiments, we randomly selected advisors according to the procedures of each experiment and sent each advisor's recommendation to a client. We informed clients ($N = 512$) that advisors had received information about the two products and had made a recommendation. Clients learned about their advisor's recommendation and then made a choice between the two products. Clients received no information about the products other than their advisor's recommendation. Overall, 82% of clients followed the advisor's recommendation.

2.4.1 Additional measures

After the recommendation stage, we elicited advisors' beliefs about the quality of Product B and some demographics. In the Choice and ChoiceStakes experiments, for participants on AMT we also collected a measure of advisors' concern for the clients' payoffs. We did not include this measure in the data collected with professionals.

Beliefs. We elicited advisors' beliefs about the likelihood that the quality of Product B was low by asking advisors i) to choose one of ten options, where Option 1 ranged between 0% and 10% and Option 10 ranged between 91% and 100%, and ii) to indicate the exact likelihood by entering a number from 0 to 100. The first measure was incentivized: Advisors received \$0.15 for a guess in the correct range.

Selfishness. We measured advisors' concern for their own payoff relative to that of the client (selfishness) using a multiple price list. We informed advisors that they would be asked to make a second recommendation, to a participant different than the one who received their first recommendation. Advisors were told they would need to make a series of recommendations between two products, X and Y. Product X varied across 5 different decisions. It paid \$2 with probabilities 1, 0.8, 0.6, 0.4, and 0 respectively, and \$0 otherwise. We always incentivized advisors to recommend Product Y and asked them to state their recommendation for the case in which the signal about the quality of Product Y was in conflict with their incentive (i.e., for the case in which the signal suggested that the quality of Y was low). We use this elicitation to measure the switch point between recommending the product associated with the commission and recommending the product that was in the advisee's best interest, and we standardize it to indicate advisor selfishness.[9] At the end of the experiment, we randomly selected one out of 10 advisors, randomly

[9]The distribution of selfishness does not differ across experiments (χ^2 test, $p = 0.171$).

picked one of the 5 recommendations, and showed them to a client. For this purpose, we recruited a total of 505 clients. Of these, 78% of clients followed the advisor's recommendation.

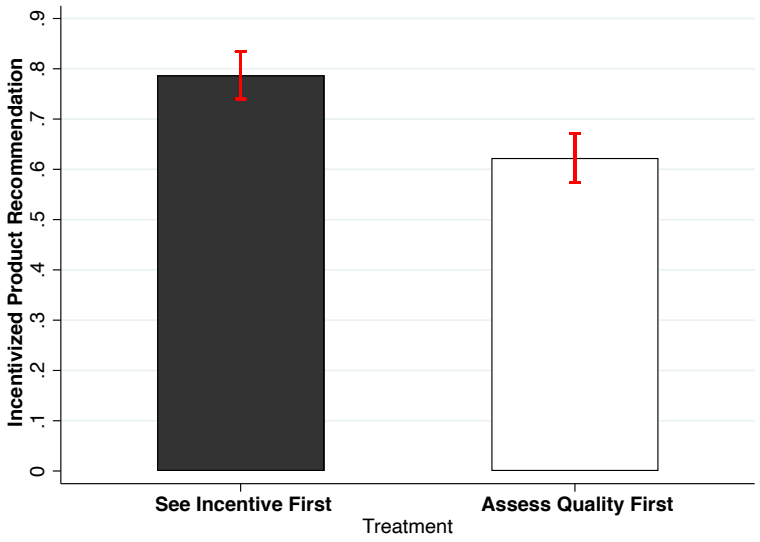
Demographics. We collected information on the participants' gender, age, their first language, ethnicity, and difficulty in understanding the instructions.

Explanations of information order preference. We added open-text explanations of choice for advisors in the second wave of data collection for the Incentives First Free treatment of the Choice experiment and to the sample of professionals. The question asked participants to explain how they made their decision to see their incentive first or the signal about product quality first.[10] Two independent raters, who were blind to advisors' choices, coded the responses of the 1,053 advisors in the convenient sample and the 713 advisors from the sample of professionals. They classified their responses into four categories, which apply to 91% of the open-ended responses. The remaining 9% consists of empty or unrelated comments. The first category was "limiting bias" and was assigned to messages that explicitly stated that the reason for their preference was to be less biased in the evaluation and to want what is best for the client. This category was meant to capture preference for commitment to accurate beliefs and moral behavior. The second category, "does not matter," captured indifference—whether advisors stated that information order did not matter. The third category, "commission," was for advisors who indicated explicitly that they cared only about their own commission. The fourth category, "other reasons," captured whether advisors indicated that gut feeling, curiosity, or other reasons guided their preference. We did not expect advisors to openly express wanting cognitive flexibility in their comments. Consistent with this, we find no such comments in the data. We allowed coders to indicate multiple categories, though this was rarely done (in less than 3% of the cases). We analyze the relationship between these categories of motives and advisor preferences in Section 5.3.

[10]The question was "When you had to decide between learning about your commission Before or After getting information about the quality of Product B [A, if the order was flipped], how did you make this decision?"

3 Does the Sequence of Information Affect Advice?

Before investigating advisors' preferences regarding the order in which they receive the signal about product quality and information about their incentive, we test whether exogenously assigning a given order of information affects advice. For the cases in which advisors face a conflict of interest (i.e., the quality signal was in conflict with their own incentive), we find that recommendations are significantly affected by the order in which information is presented to them. In the See Incentive First treatment, advisors recommend the incentivized product 79% of the time. In the Assess Quality First treatment, they recommend the incentivized product 62% of the time, as shown in Figure 2. This 17 percentage point difference is significant (Z-stat $= 2.64$, $p < 0.01$, $N = 214$.). When advisors do not face a conflict of interest, the order of information does not affect recommendations. Advisors in the See Incentive First treatment recommend the incentivized product 87% of the time, while those in the Assess Quality First treatment recommend the incentivized product 86% of the time (Z-stat $= 0.13$, $p = 0.89$, $N = 87$).



Notes: This figure shows the fraction of recommendations of the incentivized product, when there is a conflict of interest between the advisor and the client, by treatment. In the See Incentives First treatment the advisor is presented first with information about her incentive. In Assess Quality First she receives the signal about the quality of Product B first. ± 1 S.E. bars shown, N=214.

Figure 2: Recommendation of Incentivized Product, by Treatment

Therefore, the NoChoice experiment confirms that when advisors face a conflict between maximizing their own earnings and giving advice in the client's best interest, the order of information can significantly affect advice. This result is consistent with the hypothesis that advisors have larger cognitive flexibility to justify advice that is in conflict with the client's best interest when the information about their own commission is received first. Conversely, seeing information about quality first can restrict cognitive flexibility. Our data suggest that the effect of information order on recommendation is due to motivated cognition rather than being a purely cognitive bias: Altering the order of information does not affect recommendations when advisors' incentives are aligned with those of the client.

This experiment and its results set the stage for our main research questions: Which sequence of information do advisors prefer, and how does this active choice affect their recommendations?

4 Preferences for Information Order: Cognitive Flexibility or Moral Commitment?

In the Choice experiment, advisors had to indicate whether they preferred to see their incentive first or assess quality first. Advisor preferences for information order are split between cognitive flexibility and moral commitment, as shown in Figure 3. When the choice is free, 58% of advisors prefer to see the incentive information first (Incentive First Free), a fraction that is significantly different from 50% (Z-stat=7.49, p <0.001). The preference is 10 percentage points lower (Z-stat=4.71, p <0.001), but still sizable, among professionals, who prefer to see the incentive information first in 48% of the cases, a fraction that is not significantly different from 50% (Z-stat=-1.24, p = 0.2165). Conversely, between 42% and 52% of advisors choose to see the quality signal first, indicating that a substantial fraction of advisors would rather delay information about their own incentive.

When seeing the incentive first is costly, more than 2 out of 5 advisors (42%) are still willing to pay the cost (a third of their commission) to see the incentive first and have cognitive flexibility when assessing the signal. This suggests that the preference to see the incentive first, when it is free, is not driven only by indifference, as a substantial fraction shows a strict preference. There is a significant, 16-percentage-point drop relative to when it is free (Z-stat=9.20, p <0.001). One reason for this finding could be that selfish advisors who do not have self-image concerns may plan

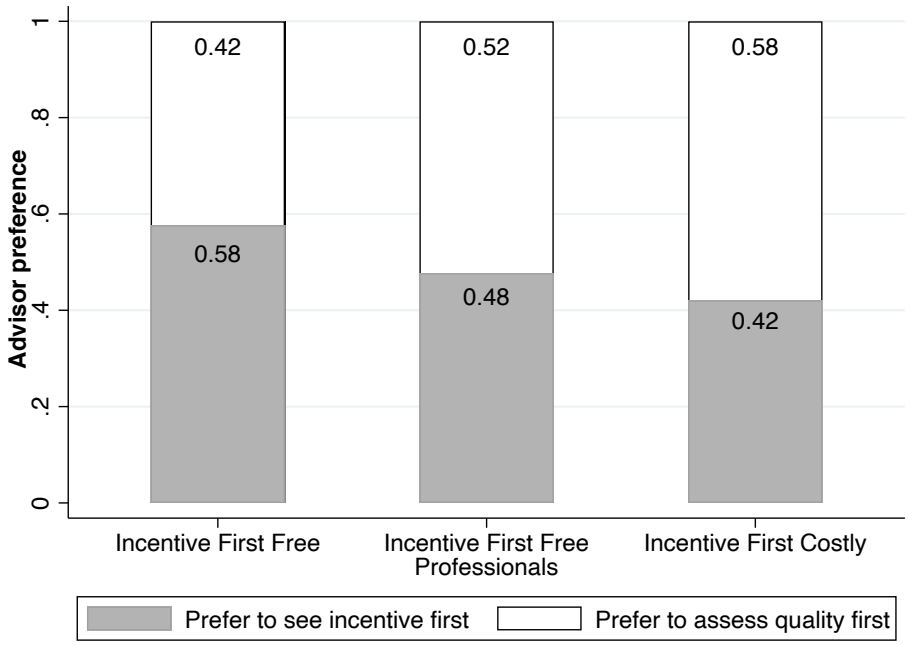


Figure 3: Preferences for Information Order

to recommend the incentivized option anyway, and therefore may not be willing to pay any positive amount to see the incentive first. Table 2 shows the determinants of the preference to see the incentive first, and column (3) tests the relationship with advisor selfishness. When the choice of information order is free, selfish advisors prefer to see the incentive first more often. This relationship weakens directionally when seeing the incentive first is costly (t-test, p=0.133), suggestive of selfish advisors without self-image concerns choosing to assess the quality first in order to avoid incurring a cost for seeing the incentive first.

Given the heterogeneity in preferences for the signal sequence, a central question is whether actively choosing a particular signal sequence affects recommendations. In the NoChoice experiment, where the information order was not willfully chosen, seeing the information about the incentive first led to higher recommendations of the incentivized product. Here, we can test whether actively pursuing cognitive flexibility by choosing to see the incentive first precludes subsequent belief distortion. That is, does the *intent* to self-deceive make self-deception ineffective? Conversely, does actively choosing to see the quality first, a form of commitment to moral behavior, also affect recommendations, making it even more salient to advisors that they are committed to providing unbiased advice to the client?

Table 2: Preference for information order

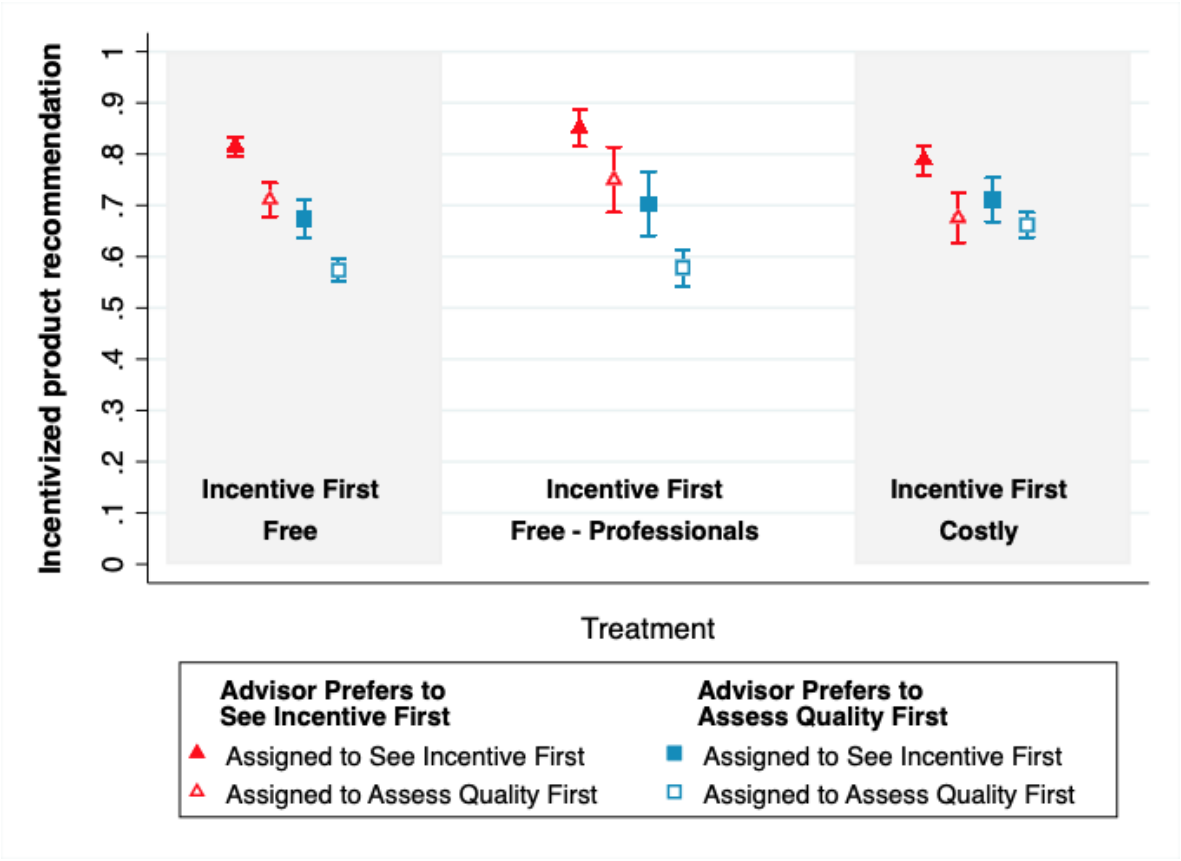
	(1)	(2)	(3)
	Prefer to see incentive first		
Incentive First Costly	-0.156***	-0.158***	-0.151***
	(0.017)	(0.017)	(0.018)
Female		-0.038**	-0.031*
		(0.015)	(0.016)
Age		-0.003***	-0.002***
		(0.001)	(0.001)
Selfishness			0.042***
			(0.010)
Incentive First Costly X Selfishness			-0.025
			(0.017)
Incentive First Free - Professionals	-0.111***	-0.111***	
	(0.024)	(0.024)	
Constant	0.577***	0.698***	0.672***
	(0.010)	(0.027)	(0.029)
Observations	4,448	4,436	3,725
R-squared	0.020	0.026	0.034

Note: This table displays the estimated coefficients from linear probability models on the preference to see the incentive first. Incentive First Costly is an indicator variable that takes value 1 if the treatment is Incentive First Costly, 0 otherwise. Selfishness is the standardized number of choices of advisors in favor of a product for the client that gives them a commission, compared to an alternative that does not. This measure was elicited at the end of the experiment, using a multiple price list with 5 decisions. The regression models in columns (2) and (3) include individual controls for the advisor's gender and age. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.1 Does Demanding Flexibility or Commitment Affect Advice?

We examine advisors' recommendation decisions conditional on their preference to see the incentive first and whether they were randomly assigned to their demanded information sequence, as shown in Figure 4. When advisors prefer to see the incentive first and are actually randomly assigned to seeing it first (leftmost triangle in each cluster in Figure 4), they recommend the incentivized product in 81.4% of the cases on average. When these advisors are instead assigned to assessing quality first, they recommend the incentivized option significantly less, in 70.1% of the cases on average (t-stat=2.79, $p = 0.005$). In Panel A of Table 3, we report coefficient estimates of a linear probability model of the advisor's decision to recommend the incentivized product for advisors who choose to see the incentive first (columns (1) and (2)) and separately for those who choose to assess quality first (columns (3)

and (4)). As shown in columns (1)-(2), the 10-percentage-point difference in recommendations does not differ significantly for professionals and is robust to whether the incentivized product is A (with advisors receiving a signal that B is of high quality) or B (with advisors receiving a signal suggesting that B is of low quality). This result shows that actively pursuing cognitive flexibility, by choosing to see the incentive first, does not limit advisors' ability to leverage that information order to their advantage and make self-serving recommendations.



Notes: This Figure shows the fraction of recommendations of the incentivized product, by advisor preference, random assignment and treatment. ± 1 S.E. bars shown.

Figure 4: Recommendations by Preferences and Random Assignment

On average, advisors who prefer to assess quality first are significantly less likely to recommend the incentivized option. When these advisors are assigned to their desired information order—seeing quality first—they recommend the incentivized product 60.7% of the time, as shown in the rightmost square in each cluster in Figure 4. However, even for advisors who prefer commitment, being randomly assigned to see the incentive first increases the percentage of advisors who recommend the

incentivized product to 69.2% of the time on average (t-stat=2.45, $p = 0.014$). Columns (3) and (4) of Panel A of Table 3 show that, for advisors who prefer to see the signal about quality first, the effect of being assigned to see the incentive first is similar to that of advisors who prefer to see the incentive first. This suggests that even if advisors prefer an information order that reduces the likelihood of making self-serving recommendations, being assigned the opposite order still leads them to "fall prey" to bias, and provide self-serving recommendations significantly more often.

Table 3 also shows that for the treatment in which seeing the incentive first is costly, advisors who prefer to assess quality first are significantly more likely to recommend the incentivized product. This is consistent with self-selection by selfish advisors: When seeing the incentive first is costly, those who plan to recommend the incentivized product switch to assessing quality first, which leads to a higher rate of recommendations of the incentivized product.

How does preferring to see the incentive first relate to recommendations? Upon indicating their preference, advisors were randomly assigned to see the incentive first. Panel B of Table 3 shows that, when assigned to their preferred information order, advisors who prefer to see the incentive first are more likely to recommend the incentivized product, with the effect being of about 20 percentage points. However, when not assigned to their preferred order, advisors who expressed a preference to see the incentive first are no more likely than advisors who indicated the opposite preference. This effect confirms that advisors who prefer to see the incentive first are significantly more likely to recommend the incentivized product than those who prefer to see the quality signal first, but only when advisors get their preferred information order.

Throughout, we find that advisors display a preference to recommend Product A over Product B. One reason for this preference is that the quality of B is uncertain, while the payoff distribution of Product A is certain. Given the payoff distributions of the two urns, rationalizing a recommendation of product A when the signal about quality favors Product B (i.e., a \$2 ball from Product B) could require less cognitive flexibility than rationalizing a recommendation of product B when the signal about product B is negative (i.e., a \$0 ball from Product B).[11] Despite the preference for Product A, the effect of seeing the incentive first is similar regardless of whether Product A or B is incentivized.

[11]The preference for Product A is not driven by an order effect: Switching the order and labels of the two products does not significantly affect recommendations (p=.515)

Table 3: Advisor Recommendations

Panel A. Effect of Assignment

	(1)	(2)	(3)	(4)
	Recommend incentivized product			
Advisor Preference:	Prefer to See Incentive First		Prefer to Assess Quality First	
Assigned to See Incentive First	0.101***	0.106***	0.087***	0.111**
	(0.025)	(0.038)	(0.027)	(0.045)
Professionals	-0.035	-0.040	-0.029	-0.034
	(0.042)	(0.069)	(0.056)	(0.060)
Professionals X Assigned to See Incentive First		0.008		0.016
		(0.069)		(0.075)
Incentive First Costly	-0.008	-0.024	0.063**	0.076**
	(0.026)	(0.054)	(0.029)	(0.033)
Incentive First Costly X Assigned to See Incentive First		0.021		-0.056
		(0.060)		(0.060)
Incentive for B	-0.152***	-0.132***	-0.215***	-0.211***
	(0.021)	(0.046)	(0.025)	(0.029)
Incentive for B X Assigned to See Incentive First		-0.026		-0.016
		(0.052)		(0.056)
Constant	0.856***	0.852***	0.784***	0.777***
	(0.042)	(0.049)	(0.047)	(0.048)
Observations	1,600	1,600	1,516	1,516
R-squared	0.063	0.063	0.067	0.068

Panel B. Preferences and Recommendations

	(1)	(2)
	Recommend incentivized product	
	If Assigned Preference	If Not Assigned Preference
Prefer to See Incentive First	0.202***	0.008
	(0.018)	(0.033)
Professionals	-0.029	-0.093
	(0.040)	(0.069)
Incentive First Costly	0.040*	0.019
	(0.022)	(0.040)
Incentive for B	-0.185***	-0.177***
	(0.019)	(0.033)
Constant	0.752***	0.911***
	(0.036)	(0.065)
Observations	2,339	777
R-squared	0.103	0.060

Notes: Panel A of this table displays the coefficient estimates of linear probability models of the advisor's decision to recommend the incentivized product when there is a conflict of interest. Columns (1) and (2) focus on advisors who prefer to see the incentive first, and columns (3)-(4) focus on those who prefer to assess quality first. Each regression model includes an indicator variable that takes value 1 for advisors who are randomly assigned to see the incentive first, and 0 otherwise. Panel B displays the coefficient estimates of regressions on the advisor's decision to recommend the incentivized product when there is a conflict of interest, by whether they were assigned to get the information according to their preference (column (1)) or not (column (2)). he regression models include individual controls for the advisor's gender and age. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In Online Appendix A, we compare the frequency of incentivized product recommendations in cases where there is no conflict of interest to the frequency of those in which there is a conflict of interest, as advisors receive signals that are in line with their desired recommendations. In line with the NoChoice experiment, our results show that being assigned to seeing the incentive first does not significantly increase recommendations of the incentivized product. The IV regressions show that when there is no conflict of interest, the effect of preferring to see the incentive first on recommendations of the incentivized product is significantly smaller, two-thirds smaller in magnitude (p <0.001).

4.2 Information Preferences and Reality Denial: Evidence of Belief Distortion

The results on recommendations demonstrate that a preference to see incentives first enables advisors to make self-serving recommendations at a higher rate. We expect this result to be due to these advisors having more cognitive flexibility to engage in reality denial, a form of self-deception whereby advisors dismiss informative signals that are in conflict with their incentive, when actually assigned to see their incentive first. To examine whether seeing the incentive first leads to more reality denial, we make use of the data on beliefs and study two outcome measures. First, we investigate whether advisors who demand flexibility are less likely to update in the Bayesian direction when they get to see the incentive first. Second, we investigate how distant these advisors' posterior beliefs about the quality of Product B are from the prior. These outcomes are based on our continuous measure of beliefs, for which advisors reported the exact likelihood that Product B was of low quality. We find similar effects in our interval measure of beliefs, which was incentivized, as shown in Online Appendix A.

In Table 4, we report the results of OLS regressions where we investigate how preferring to see the incentive first relates to the beliefs of participants who got assigned to their preferred information order (Columns 1 and 3) or not (columns 2 and 4). Column (1) of Table 4 shows that, for those participants who were randomly assigned to their preferred information order, advisors who prefer to see the incentive first are 7.9 percentage points less likely to update in the Bayesian direction; this is not the case for those who were not assigned to their preferred order, as shown in column (2). Column (3) reveals that advisors who were assigned to their preferred order and preferred to see their incentive first, were significantly closer to the prior

of 50%. On average, advisors' beliefs move by about 10 percentage points toward the Bayesian posterior, which is 75% in case the signal is good for Product B, and 33% in case the signal is bad, if they are assigned their preference and they prefer to see the signal first. Their behavior hence displays conservatism, in line with existing literature on asymmetric updating in response to bad news (e.g., Möbius et al., 2014). When advisors prefer to see the incentive first, their beliefs are 2 percentage points closer to the prior (approximately 20%), in line with reality denial. As with recommendations, there is no difference in beliefs depending on advisor's preferences, if they are not assigned their preference.

Table 4: Beliefs about Quality of Product B

	(1)	(2)	(3)	(4)
	Update in Bayesian Direction=1		Distance from Prior	
	Assigned Preference	Not Assigned Preference	Assigned Preference	Not Assigned Preference
Prefer to See Incentive First	-0.079***	0.002	-2.185***	-0.060
	(0.021)	(0.036)	(0.715)	(1.272)
Professionals	0.036	-0.040	2.473	-0.216
	(-0.020)	(0.083)	(1.640)	(2.950)
Incentive First Costly	-0.020	-0.010	-1.198	-0.091
	(0.025)	(0.044)	(0.865)	(1.489)
Incentive for B	-0.013	0.048	-0.960	0.228
	(0.021)	(0.036)	(0.721)	(1.277)
Constant	0.576***	0.435***	10.056***	5.576**
	(0.042)	(0.073)	(1.425)	(2.438)
Observations	2,339	777	2339	777
R-squared	0.014	0.007	0.008	0.004

Notes: This table displays the coefficient estimates of OLS regressions on the advisor's beliefs about the likelihood that Product B is of low quality when there is a conflict of interest. Columns (1) and (3) focus on the cases in which advisors were assigned their preferences whereas columns (2) and (4) focus on the cases in which advisors were not assigne their preferred information order. The dependent variable (DV) in columns (1) and (2) is an indicator that takes value 1 if the advisor's belief that B is of low quality moves away from 50% (the prior) toward the Bayesian posterior. The DV in columns (3) and (4) is the distance between the advisor's belief and the prior of 50%. This distance, between 0 and 100, is positive if the advisor updates in the Bayesian direction, and negative otherwise. The regression models include individual controls for the advisor's gender and age. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The results in Table 4 explore whether beliefs differ depending on whether the incentive was for Product A or Product B. We do not find a significant difference in beliefs, though advisors' beliefs are somewhat less distant from the prior when the incentive is for Product B. Dismissing a positive signal about Product B when the incentive was for Product A may have been easier for advisors than dismissing a

negative signal about Product B when they were incentivized to recommend Product B.

4.3 Advisors' Explanations of Their Preferences

As described above, for a subset of advisors in the Incentive First treatment and for professionals, we elicited explanations of their preference to see their incentive first or the quality signal first, using an open ended question at the end of the experiment. The two raters agreed in over 82% of their classifications, leading to an interrater agreement κ of 0.76. We average their ratings to examine how advisors' explanations vary with their preference of information order (detailed results provided in Online Appendix A). Advisors rarely report that they are indifferent between seeing the incentive first or assessing quality first ("does not matter" category is assigned to 8% of the comments), which suggests that indifference is not a main driver of choices. The results indicate that the reasons (categories) reported by advisors who preferred to see the quality signal first differ significantly from those of advisors who preferred to see their incentive first (χ^2-stat = 461, p <0.001). Consistent with the interpretation that seeing the signal quality first acts as a form of moral commitment, in 47% of the cases (42% for AMT participants and 54% for professionals) advisors who select to see the signal of quality first directly report doing so to limit bias in their evaluation. By contrast, only 6% (5% for AMT and 7% for professionals) of those who select to see the incentive first report such a motivation. Advisors who prefer to see the commission first report to be interested in the commission (in 36% of the cases both for AMT and for professionals), and to be driven by other reasons such as gut feelings or curiosity (in 55% of the cases for AMT and professionals). This analysis provides further evidence consistent with the interpretation that many advisors anticipated that seeing the quality signal first limits cognitive flexibility (and potential bias), and preferred the opportunity to commit to accurate and therefore moral judgment.

5 Do Advisors Anticipate the Effects of Choosing Cognitive Flexibility or Moral Commitment?

In two additional experiments, we further investigate whether individuals anticipate the effect of choosing information sequences on behavior.

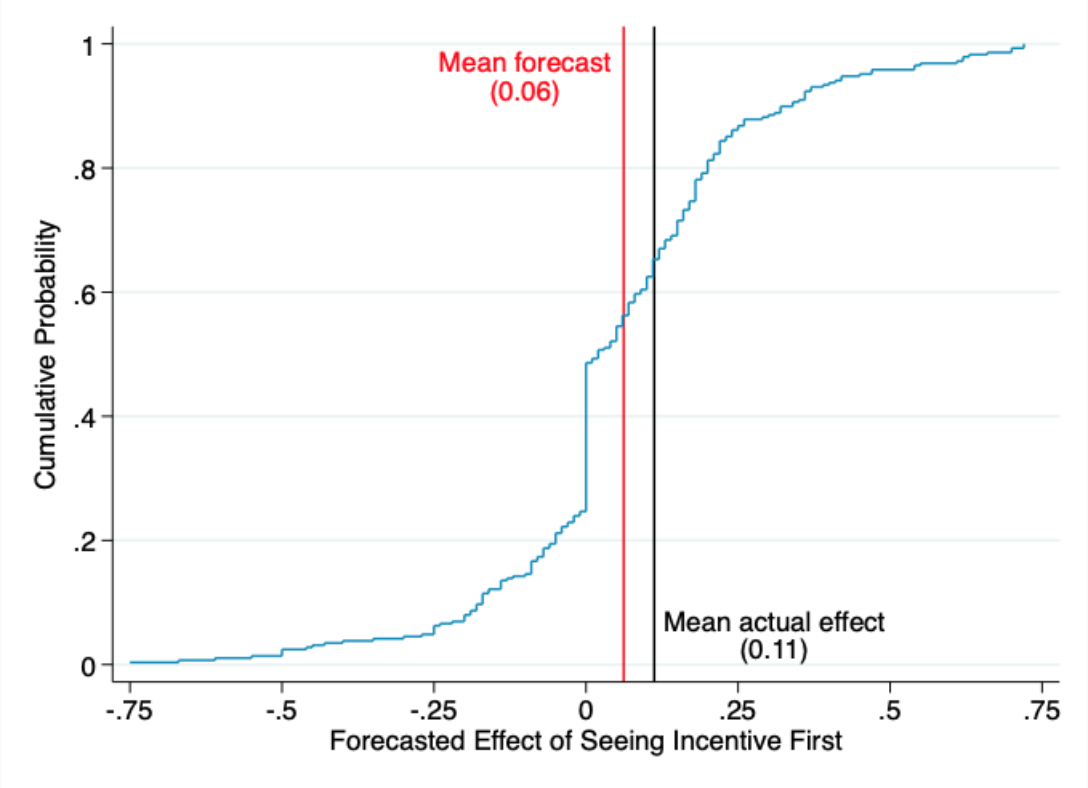
5.1 Predictions: Do Forecasters Anticipate the Effects of Information Sequences on Advisors' Bias?

In order to interpret advisors' preferences to see their incentive first or, on the contrary, assess quality first, as evidence that individuals actively pursue or constrain cognitive flexibility, it is important to test whether individuals anticipate that the order of information will affect their recommendations. To investigate this question, we turn to the Prediction experiment, in which a group of forecasters predicted the difference in recommendations between the two information orders for the case in which seeing the incentive first is costly. Figure 5 shows the cumulative distribution function of forecasts, as well as the average predicted effect and the average actual effects of seeing the incentive first. The predicted effect of seeing the incentive first—relative to seeing quality first—is 6.2 percentage points (SE=0.12, $N = 288$). This is significantly different from zero (p <0.001). It is not significantly different from the actual effect of 11.2 percentage points ($p = 0.3678$), which we documented in the Choice experiment. Whether there is an incentive to recommend Product A or Product B does not influence the predicted effect of information order on recommendations ($p = 0.7922$), in line with the actual recommendations. As shown in Figure 5, the majority of participants expect a positive effect of seeing the incentive first (51.4%), while 24.0% predict no effect and 24.6% predict a negative effect.

This experiment therefore provides some evidence that individuals evaluating the task of advisors can anticipate the effects of seeing the incentive first, although on average they may somewhat underestimate the magnitude of those effects. This result is consistent with the interpretation that the choice to see the incentive first or assess quality first is at least in part driven by the anticipated effect of this information order on recommendations.

5.2 Advisors' Preferences and Incentives

In the ChoiceStakes experiment, we test whether, as hypothesized, advisors demand to see the incentive first because of the desire to have cognitive flexibility, earning incentives while convincing themselves that their desired recommendation is also the best recommendation for the client. If the gains from recommending the incentivized option decrease, advisors have a smaller incentive to distort their beliefs, making the demand for cognitive flexibility (seeing the incentive first) less desirable. The ChoiceStakes experiment varies the size of the advisors' incentive, to test whether

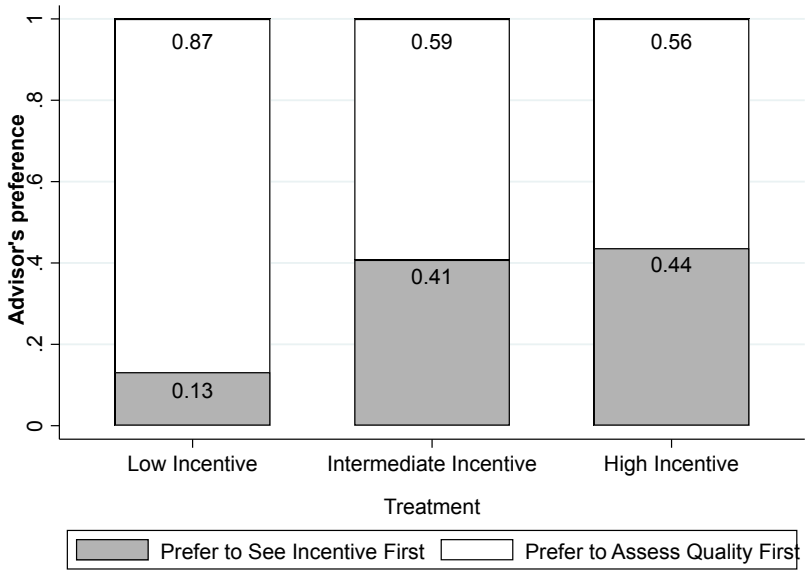


Notes: This figure displays the distribution of forecasts regarding the effect of seeing the incentive first on recommendations of the incentivized product in the Choice experiment, for advisors who prefer to see the incentive first when seeing it is costly, the average forecast (from the Predictions experiment) and the average actual effect (from the Choice experiment).

Figure 5: Predicted and Actual Effect of Seeing the Incentive First on Recommendations

the preference to see the incentive first is driven by the potential gains from self-deception. Figure 6 shows the advisors' preference to see the incentive first. In the Intermediate incentive treatment, 41% of advisors prefer to see the incentive first, replicating our finding in the Incentives First Costly treatment of Choice Experiment (where it was 42%). This fraction decreases significantly in the Low Incentive treatment, to 13% (Z-stat=9.79, p <0.001). In the High Incentive treatment, advisor's preference to see the incentive first increases by only 3 percentage points, to 44% (Z-stat=0.92, p =0.3575), despite the fact that the commission is doubled. These results are confirmed in regression analyses in Online Appendix A.

We conduct exploratory analyses on advisors' recommendations of the incentivized product in each treatment, shown in Table 5. We pool all treatments together, and test whether advisors who prefer to see the incentive first are more



Notes: This figure shows the fraction of advisors who prefer to see their incentive first. In the Low Incentive treatment the commission for learning before is \$0.01, in the Intermediate Incentive treatment it is \$0.15, and in the High Incentive treatment it is \$0.30. Seeing the incentive first costs \$0.05 in all treatments, as in the Incentive First Costly treatment of the Choice experiment.

Figure 6: Advisor's Preference to See Incentive First, by Treatment

likely to recommend the incentivized product, as in the Choice experiment. We do not examine the differential effects of preferences for flexibility or commitment by incentive, since the sample sizes become small, especially in the LowIncentive treatment, and power is then very limited. Overall, seeing the incentive first increases the likelihood that the advisor recommends the incentivized product. In line with standard incentive effects, the likelihood of recommending the incentivized product drops significantly in the Low Incentive treatment, while it marginally increases in the High Incentive treatment.

Beliefs about the quality of Product B exhibit qualitatively similar patterns to those found in the Choice experiment (detailed results in Online Appendix A). In all treatments, the distance between the advisors' posterior and prior is smaller for advisors who prefer to see the incentive first.

This experiment shows that advisors' preferences to see the incentive first in the Choice experiment does not appear to be explained by alternative explanations, such as curiosity, which can be considered independent of the stakes of the advisor. It also reveals that advisors' preferences to see the incentive first exhibit a concave

Table 5: Advisor Recommendations (IV Regression)

	(1)	(2)
	Recommend Incentivized Product	
Prefer to See Incentive First	0.268***	0.271***
	(0.067)	(0.067)
Incentive for B	-0.124***	-0.125***
	(0.030)	(0.030)
Selfishness		0.047***
		(0.015)
Low Incentive	-0.081**	-0.083**
	(0.041)	(0.041)
High Incentive	0.066*	0.064*
	(0.036)	(0.036)
Constant	0.577***	0.590***
	(0.039)	(0.049)
Controls	No	Yes
Observations	1,033	1,031
R-squared	0.053	0.064

Notes: This table displays the coefficient estimates of IV regressions on the advisor's decision to recommend the incentivized product when there is a conflict of interest. The advisor's preference to see the incentive first is instrumented by the random assignment to see the incentive first. Selfishness is the standardized number of choices of advisors in favor of a product for the client that gives them a commission, compared to an alternative that does not. This measure was elicited at the end of the experiment, using a multiple price list with 5 decisions. The regression models include individual controls for the advisor's gender and age. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

shape. When doubling the commission of the advisor, the preference to see the incentive first increases by only 3 percentage points, less than 10 percent.

6 Conclusion

A large body of research has shown that individuals care about upholding a positive identity, and may distort informative signals to preserve their self-image. However, the ability to engage in belief distortion when confronted with unpleasant information is subject to cognitive constraints: Altering beliefs is easier in situations where advisors have cognitive flexibility. We ask: If given the choice, do individuals prefer to limit cognitive flexibility, committing themselves to unbiased judgment, or instead seek out flexibility, in order to have larger scope for justifying moral transgressions?

Our goal is to provide the first evidence of individuals' sophistication about the malleability of beliefs and cognitive constraints to ex-post belief distortion in response to potentially inconvenient information. Our experiments examine individuals' preferences for cognitive flexibility or commitment to more accurate beliefs in a series of experiments in which advisors face a potential conflict of interest and can choose the order with which they receive a sequence of signals. One sequence first gives information about what is best for the advisors and only subsequently provides information about what is best for the client (in terms of product quality). The other sequence reverses the order. Seeing the incentive first can make this information more salient, providing more cognitive flexibility to dismiss subsequent inconvenient signals and rationalize self-serving behavior. Conversely, seeing the signal about quality first constrains cognitive flexibility, committing individuals to more accurate beliefs and more moral recommendations.

We find that the preferences of advisors are heterogeneous. Although a considerable fraction of advisors prefers commitment, a sizable fraction of individuals (more than 40%) seeks out cognitive flexibility, by asking to see their incentive before making quality assessments, even when receiving information in this order is costly. Individuals who actively seek cognitive flexibility are able to still exhibit significant bias in their recommendations and to hold significantly more biased beliefs, suggesting that *intending* to distort beliefs does not preclude belief distortion. Our data from forecasters confirms that the effects of seeing the incentives first are anticipated; and we find that preferences for cognitive flexibility respond to advisors' incentives to rationalize self-serving behavior. Taken together, these results provide empirical evidence that that individuals anticipate the cognitive constraints to belief distortion, suggesting some level of sophistication about their ability to distort their beliefs. Whereas this work focuses on the moral domain, future work could investigate whether, in other domains, individuals similarly anticipate the effect of

altering the sequence of information in order to preserve motivated beliefs about, for example, their intelligence or skills, in spite of threatening feedback.

The results of our experiments also show that advisors who preferred commitment—wanting to first assess the quality of the product—but were assigned to first learn about their incentives, were more likely to provide biased recommendations than advisors who received information in their preferred order. This finding suggests that actively wanting to commit to unbiased (and moral) judgment may not be enough to prevent biased recommendations when the environment in which advisors make decisions is structured in a way that amplifies cognitive flexibility (see Epley and Tannenbaum, 2017).

From a policy perspective, our results suggest that in the context of fiduciary relationships with conflict of interest, presenting fiduciaries with all of the relevant information about what is best for a third party's may not be enough to prevent self-serving behavior. To ensure ethical advice, it is also crucial to consider how information is presented and who dictates the information design. Our findings suggest that experts who face potential conflicts of interest may be able to anticipate the effects of small changes in the way information is presented, and more specifically how altering the order with which different pieces of information are processed can either limit or enable self-serving belief distortion. If so, these experts may design institutions and codes of conduct that maximize cognitive flexibility, enhancing their scope to behave self-servingly while preserving a self-image of ethicality.

Reference List

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

Amir, O., Rand, D.G., and Gal, Y.K. (2012). Economic games on the Internet: The effect of $1 stakes. *PloS ONE*, 7(2): e31461.

Babcock, L., Loewenstein, G., Issacharoff, S., and Camerer, C. (1995). Biased judgments of fairness in bargaining. *The American Economic Review*, 85(5), 1337–1343.

Bénabou, R. (2013). Groupthink: Collective delusions in organizations and markets.

Review of Economic Studies, 80(2), 429–462.

Bénabou, R. (2015). The economics of motivated beliefs. Jean-Jaques Laffont Lecture, *Revue d'Économie Politique*, 125(5), 665–85.

Bénabou, R., Falk, A., and Tirole, J. (2018). Narratives, Imperatives and Moral Reasoning. NBER Working Paper #24798.

Bénabou, R., and Tirole, J. (2002). Self-confidence and personal motivation. *Quarterly Journal of Economics*, 117(3), 871–915.

Bénabou, R., and Tirole, J. (2006). Incentives and prosocial behavior. *The American Economic Review*, 96(5), 1652–1678.

Bénabou, R., and Tirole, J. (2011). Identity, morals and taboos: Beliefs as assets. *Quarterly Journal of Economics*, 126(2), 805–55.

Bénabou, R., and Tirole, J. (2016). Mindful Economics: The production, consumption and value of beliefs. *Journal of Economic Literature*, 30(3), 141–64.

Benjamin, D.J. (2019). Errors in probabilistic reasoning and judgment biases (Chapter 2). *Handbook of Behavioral Economics: Applications and Foundations* Vol. 2, 69–186.

Bermúdez, J.L. (2000). Self-deception, intentions, and contradictory beliefs. *Analysis*, 60(4), 309–319.

Bodner, R., and Prelec, D. (2003). Self-signaling and diagnostic utility in everyday decision making. *The Psychology of Economic Decisions*, 1(105), 26.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2012). Salience theory of choice under risk. *Quarterly Journal of Economics*, 127(3), 1243–1285.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2013). Salience and consumer choice. *Journal of Political Economy*, 121(5), 803–843.

Brunnermeier, M.K., and Parker, J.A. (2005). Optimal expectations. *American Economic Review*, 95(4), 1092–1118.

Carlson, R.W., Marechal, M., Oud, B., Fehr, E., and Crockett, M. (2020). Motivated misremembering of selfish decisions. *Nature Communications*, 11(1), 1–11.

Chetty, R., Looney, A., and Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review*, 99(4), 1145–77.

Cohn, A., Marechal, M.A., Tannenbaum, D., and Zund, C.L. (2019). Civic honesty around the globe. *Science*, 365(6448), 70–73.

Crawford, V., and Sobel, J. (1982). Strategic information transmission. *Econometrica*, 50(6), 1431–1451.

Dana, J., Weber, R. A., and Kuang, J.X. (2007). Exploiting moral wriggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33(1), 67–80.

Darby, M.R., and Karni, E. (1973). Free competition and the optimal amount of fraud. *The Journal of Law and Economics*, 16(1), 67–88.

DellaVigna, S., and Pope, D. (2018). Predicting experimental results: Who knows what? *Journal of Political Economy*, 126(6), 2410–2456.

DellaVigna, S., Pope, D., and Vivalt, E. (2019). Predicting science to improve science. *Science*, 366(6464), 428–429.

Di Tella, R., Perez-Truglia, R., Babino, A., and Sigman, M. (2015). Conveniently upset: Avoiding altruism by distorting beliefs about others' altruism. *American Economic Review*, 105(11), 3416–3442.

Ditto, P.H., and Lopez, D.F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusion. *Journal of Personality and Social Psychology*, 63(4), 568–584.

Eil, D. and Rao, J.M. (2011). The good news-bad news effect: Asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, 3(2), 114–38.

Epley, N., and Gilovich, T. (2016). The mechanics of motivated reasoning. *Journal of Economic Perspectives*, 30(3), 133–40.

Epley, N., and Tannenbaum, D. (2017). Treating ethics as a design problem. *Behavioral Science and Policy*, 3(2), 72–84.

Exley, C.L. (2015). Excusing selfishness in charitable giving: The role of risk. *The Review of Economic Studies*, 83(2), 587–628.

Exley, C.L., and Kessler, J.B. (2019). Motivated errors. NBER Working Paper #26595.

Falk, A., Neuber, T., and Szech, N. (2020). Diffusion of being pivotal and immoral outcomes. *The Review of Economic Studies*, forthcoming.

Fehr. E., and Rangel, A. (2011). Neuroeconomic foundations of economic choice – Recent advances. *Journal of Economic Perspectives*, 25(4), 3–30.

Gabaix, X., Laibson, D., Moloche, G., and Weinberg S. (2006). Costly information acquisition: Experimental analysis of a boundedly rational model. *American Economic Review*, 96(4), 1043–1068.

Ganguly, A. and Tasoff, J. (2017). Fantasy and dread: The demand for information and the consumption utility of the future. *Management Science*, 63(12), 4037–4060.

Gino, F., Norton M., and Weber, R. (2016) Motivated Bayesians: Feeling moral while acting egoistically. *Journal of Economic Perspectives*, 30(3), 189–212.

Gneezy, U., Saccardo S., and van Veldhuizen R. (2018). Bribery: Behavioral drivers of distorted decisions. *Journal of the European Economic Association*, 17(3), 917–946

Gneezy, U., Saccardo S., Serra-Garcia, M., and van Veldhuizen R. (2020). Bribing the self. *Games and Economic Behavior*, 120, 917–946.

Goldin, C. and Rouse, C. (2000). Orchestrating impartiality: The impact of "blind" auditions on female musicians. *American Economic Review*, 90(4), 715–741.

Golman, R., Hagmann, D., and Loewenstein, G. (2016). Information avoidance. *Journal of Economic Literature*, 55(1), 96–135.

Golman, R., Molnar, A., Loewenstein, G., and Saccardo, S. (2019). The demand of, and avoidance of information. *Mimeo.*

Grossman, Z. (2014). Strategic ignorance and the robustness of social preferences. *Management Science*, 60(11), 2659–2665.

Grossman, Z., and van Der Weele. J.J. (2017). Self-image and willful ignorance in social decisions. *Journal of the European Economic Association*, 15(1), 173–217.

Haisley, E.C., and Weber, R.A. (2010). Self-serving interpretations of ambiguity in other-regarding behavior. *Games and Economic Behavior*, 68(2), 614–625.

Hsee, C.K. (1996). Elastic justification: How unjustifiable factors influence judgments. *Organizational Behavior and Human Decision Processes*, 66(1), 122–129.

Huffman, D., Raymond, C., and Shvets, J. (2020). Persistent overconfidence and biased memory: Evidence from managers. Working paper.

Kahan, D. (2013). Ideology, motivated reasoning, and cognitive reflection: An experimental study. *Nature*, 488, 255.

Konow, J. (2000). Fair shares: Accountability and cognitive dissonance in allocation decisions. *The American Economic Review*, 90(4), 1072–1091.

Kouchaki, M., and Gino, F. (2016). Memories of unethical actions become obfuscated over time. *Proceedings of the National Academy of Sciences*, 113(22), 6166–6171.

Köszegi, B. (2006). Ego utility, overconfidence, and task choice. *Journal of the European Economic Association* 4(4), 673–707.

Köszegi, B., and Szeidl, A. (2013). A model of focusing in economic choice. *Quarterly Journal of Economics* 128(1), 53–104.

Kunda, Z. (1990). The Case for Motivated Reasoning. *Psychological Bulletin* 108(3), 480–98.

Larson, T., and Capra, C.M. (2009). Exploiting moral wiggle room: Illusory preference for fairness? A comment. *Judgment and Decision Making*, 4(6), 467.

Litman, L., Robinson, J., and Abberbock, T. (2016). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, 1–10.

Malmendier, U., and Tate, G. (2005). CEO overconfidence and corporate investment. *Journal of Finance*, 60(6), 2661–2700.

Malmendier, U., and Tate, G. (2008). Who makes acquisitions? CEO overconfidence and the market's reaction. *Journal of Financial Economics*, 89(1), 20–43.

Malmendier, U., and Schmidt, K. (2017). You owe me. *American Economic Review*, 107(2), 493–526.

Mele, A. (1987). *Irrationality: An Essay on Akrasia, Self-Deception, Self-Control.* Oxford: Oxford University Press.

Mele, A. (2001). *Self-Deception Unmasked.* Princeton: Princeton University Press.

Mijovic-Prelec, D., and Prelec, D. (2010). Self-deception as self-signaling: A Model and experimental evidence. *Philosophical Transactions of the Royal Society B*, 365,227–240.

Moore, D. A., Tanlu, L., and Bazerman, M. H. (2010). Conflict of interest and the intrusion of bias. *Judgment and Decision Making*, 5(1), 37.

Möbius, M., Niederle, M., Niehaus, P. and Rosenblat, T. (2013). Managing self-confidence. Mimeo.

Oster, E., Shoulson, I., and Dorsey, E. (2013). Optimal expectations and limited medical testing: Evidence from Huntington disease. *American Economic Review*, 103(2), 804–30.

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

Paolacci, G., Chandler, J., and Ipeirotis, P.G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making* 5(5), 411–419.

Pitchik, C., and Schotter, A. 1987. Honesty in a model of strategic information transmission. *The American Economic Review*, 77(5), 1032–1036.

Quattrone, G.A., and Tversky, A. (1984). Causal versus diagnostic contingencies: On self-deception and on the voter's illusion. *Journal of Personality and Social Psychology*, 46(2), 237.

Saucet, C., and Villeval, M.C. (2019). Motivated memory in dictator games. *Games and Economic Behavior*, 117, 250–275.

Schwardmann, P., Tripodi, E. and van der Weele, J.J. (2019). Self-Persuasion: Evidence from Field Experiments at Two International Debating Competitions. *mimeo*

Schwartzstein, J. (2014). Selective attention and learning. *Journal of the European Economic Association* 12(6), 1423–1452.

Serra-Garcia, M., and Szech, N., (2019). The (in) elasticity of moral ignorance. *CESifo Working Paper No. 7555.*

Shalvi, S., Dana, J., Handgraaf, M. J., and De Dreu, C. K. (2011). Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes*, 115(2), 181–190.

Shalvi, S., Gino, F., Barkan, R., and Ayal, S. (2015). Self-serving justifications: Doing wrong and feeling moral. *Current Directions in Psychological Science*, 24(2), 125–130.

Sharot, T., Korn, C.W., and Dolan, R.J., (2011). How unrealistic optimism is maintained in the face of reality. *Nature Neuroscience*, 14(11), 1475–1479

Sicherman, N., Loewenstein, G., Seppi, D.J., and Utkus, S.P. (2016). Financial Attention. *Review of Financial Studies*, 29(4), 863–897.

Sloman, S.A., Fernbach, P.M., and Hagmayer, Y. (2010). Self-deception requires vagueness, *Cognition*, 115(2), 268–281.

Sobel, J., 2020. Lying and deception in games. *Journal of Political Economy*, 128(3), 907–947.

Tasoff, J., and Madarasz, K. (2009). A model of attention and anticipation. Working paper.

Trivers, R. (2011). *The Folly of Fools: The Logic of Deceit and Self-Deception in Human Life.* Basic Books.

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

Zimmerman, F. (2018). The dynamics of motivated beliefs. *American Economic Review*, 110(2), 337–61.

For Online Publication

APPENDIX A: Additional Results

A.1. Balance Check & Recruitment Details

Table A.1. displays the average age and share of female participants in each experiment and treatment. Within each experiment that included multiple treatments, we test for balance in gender and age. We do not find a significant difference across treatments, except for the case of age in the NoChoice Experiment. We control through age and gender throughout in the analysis.

Table A.1. Balance Check

	Age (Mean)	Female (%)	N
NoChoice Experiment			
See Incentive First Treatment	35.4	0.48	153
Assess Quality First Treatment	38.7	0.53	148
H_0: no treatment difference, p-value	0.02	0.33	
Choice Experiment			
Incentive First Free	37.7	0.52	2377
Incentive First Free - Professionals	36.7	0.53	713
Incentive First Costly	37.1	0.51	1358
H_0: no treatment difference, p-value	0.36	0.68	
Prediction Experiment	36.2	0.50	288
ChoiceStakes Experiment			
Low Treatment	37.2	0.56	486
Intermediate Treatment	36.7	0.57	515
High Treatment	38.3	0.56	488
H_0: no treatment difference, p-value			
Low vs. Intermediate	0.69	0.77	
Low vs. High	0.36	0.94	
Intermediate vs. High	0.18	0.50	

Recruitment Procedures

The experiments were conducted on Amazon Mechanical Turk (AMT), except for the study with professionals, conducted on Prolific and Cloudresearch. We pre-registered the design, sample sizes, exclusion criteria, and analyses of all AMT experiments on aspredicted.org. We recruited participants in the role of advisors to a 5 minutes study on decision-making and compensated them with $0.50 for completing the study and providing a recommendation to a participant in the role of client. Participants had to be located in the US and have an approval rating higher than 95%. Participants were presented with several understanding questions while reading the instructions. We included one question that participants had to answer correctly in order to continue in the study. Those who failed to answer it correctly, were disqualified from participation. As pre-registered, we focus the analysis on participants who provided consistent responses in all the tasks and passed the attention check question.

A.2. Additional Results: Choice Experiment

Table A.2. Advisor Recommendations in the Absence of a Conflict of Interest

	(1)	(2)	(3)	(4)
	Recommend incentivized product			
Advisor Preference:	Prefer to See Incentive First		Prefer to Assess Quality First	
Assigned to See Incentive First	0.033	0.016	0.041	0.057
	(0.032)	(0.040)	(0.033)	(0.058)
Professionals	0.003	0.060	-0.157**	-0.159**
	(0.054)	(0.081)	(0.066)	(0.068)
Professionals X Assigned to See Incentive First		-0.076		0.021
		(0.083)		(0.105)
Incentive First Costly	-0.008	0.015	0.074**	0.062
	(0.032)	(0.076)	(0.036)	(0.043)
Incentive First Costly X Assigned to See Incentive First		-0.028		0.047
		(0.082)		(0.069)
Incentive for B	-0.137***	-0.178***	-0.116***	-0.103***
	(0.022)	(0.049)	(0.029)	(0.036)
Incentive for B X Assigned to See Incentive First		0.052		-0.054
		(0.055)		(0.060)
Age	-0.001	-0.001	-0.004***	-0.004***
	(0.001)	(0.001)	(0.001)	(0.001)
Female	-0.046*	-0.047*	-0.028	-0.025
	(0.025)	(0.025)	(0.030)	(0.030)
Constant	1.015***	1.028***	1.010***	1.007***
	(0.056)	(0.056)	(0.062)	(0.064)
Observations	679	679	641	641
R-squared	0.048	0.050	0.056	0.057

Note: This table displays advisor recommendations when there is no conflict of interest with the client. The variables are the same as in Table 2. The regression models include individual controls for the advisor's gender and age. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.3. Effect of Preference with and without Conflict of Interest (IV Regressions)

	(1)	(2)
	Recommend incentivized product	
Prefer to See Incentive First	0.300***	0.316***
	(0.032)	(0.044)
No Conflict of Interest	0.275***	0.276***
	(0.030)	(0.031)
No Conflict of Interest X Prefer to See Incentive First	-0.200***	-0.201***
	(0.050)	(0.051)
Professionals	-0.041	-0.057
	(0.028)	(0.046)
Professionals X Prefer to See Incentive First		0.039
		(0.072)
Incentive First Costly	0.053***	0.098***
	(0.016)	(0.034)
Incentive First Costly X Prefer to See Incentive First		-0.100*
		(0.058)
Incentive for B	-0.165***	-0.174***
	(0.013)	(0.031)
Incentive for B X Prefer to See Incentive First		0.018
		(0.051)
Age	-0.002***	-0.002***
	(0.001)	(0.001)
Female	0.011	0.011
	(0.013)	(0.013)
Constant	0.689***	0.679***
	(0.034)	(0.038)
Observations	4,436	4,436
R-squared	0.069	0.072

Note: This table displays advisor recommendations both when there is and when there is no conflict of interest with the client. The variables are the same as in Table 2. The regression models include individual controls for the advisor's gender and age. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4. Choice Experiment: Incentivized Belief Question Correct

	(1) Correct Belief Bin=1	(2) Correct Belief Bin=1
Prefer to See Incentive First	-0.069***	-0.069***
	(0.025)	(0.025)
Incentive First Costly	-0.036**	-0.040**
	(0.015)	(0.019)
Incentive for B		0.003
		(0.015)
Incentive for B X Incentive First Costly		0.012
		(0.027)
Professionals	0.157***	0.157***
	(0.027)	(0.027)
Constant	0.233***	0.231***
	(0.028)	(0.029)
Observations	3,116	3,116
R-squared	0.025	0.025

Note: This table displays IV regressions on advisor beliefs based on the belief question that asked them to choose 1 out of 10 possible bins regarding the likelihood that the quality of Product B was low. This question was incentivized. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5. Advisors' Explanations: Detailed Results

		Advisors' Explanations of Preference (Categories)			
		Limiting Bias	Indifference	Commission	Other reasons
Sample: All (N=1,567)					
Prefer to:	Assess Quality First	47.4%	10.8%	7.8%	37.3%
	See Incentive First	5.8%	7.1%	36.3%	55.3%
Sample: AMT (N=907)					
Prefer to:	Assess Quality First	41.6%	11.5%	10.6%	39.4%
	See Incentive First	5.0%	7.4%	36.3%	55.3%
Sample: Professionals (N=660)					
Prefer to:	Assess Quality First	53.7%	10.0%	4.8%	34.9%
	See Incentive First	7.2%	6.6%	36.4%	55.2%

Note: This table displays the fraction of advisors whose explanation to see their incentive first or assess quality first was classified into each category. This classification excludes answers that were blank or unrelated to the choice.

A.3. ChoiceStakes Experiment: Additional Results

ChoiceStakes Experiment: Preference to See Incentive First

	Advisor Prefers to See Incentive First	
	(1)	(2)
Low Incentive Treatment	-0.276***	-0.277***
	(0.028)	(0.028)
High Incentive Treatment	0.029	0.025
	(0.028)	(0.028)
Selfishness		0.021*
		(0.012)
Constant	0.408***	0.426***
	(0.020)	(0.032)
Controls	No	Yes
Observations	1489	1487

Notes: This table displays the estimated coefficients from linear probability models on the preference of the advisor to see the incentive first. Low Incentive Treatment is an indicator variable that takes value 1 if the treatment is Low Incentive, 0 otherwise. High Incentive Treatment is an indicator variable that takes value 1 if the treatment is High Incentive, 0 otherwise. Selfishness is the standardized number of choices of advisors in favor of a product for the client that gives them a commission, compared to an alternative that does not. This measure was elicited at the end of the experiment, using a multiple price list with 5 decisions. The regression models include individual controls for the advisor's gender and age. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

ChoiceStakes Experiment: Beliefs

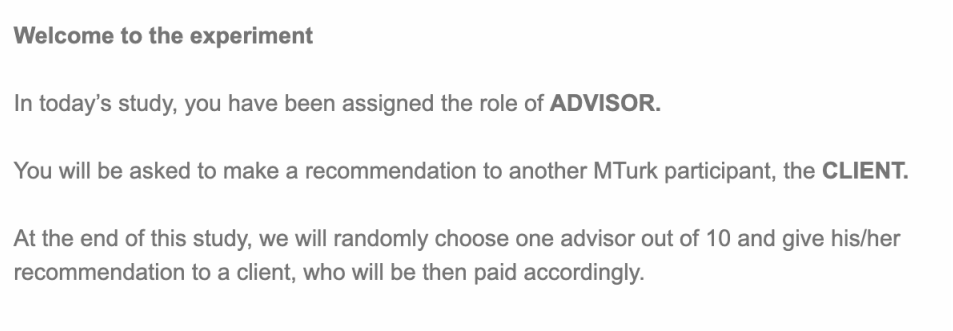
	(1)	(2)	(3)	(4)
	Update in Bayesian Direction		Distance from Prior	
Prefer to See Incentive First	-0.071	-0.072	-4.493*	-4.639*
	(0.070)	(0.070)	(2.414)	(2.410)
Incentive First Costly Treatment	-0.005	-0.006	0.128	0.161
	(0.031)	(0.031)	(1.078)	(1.073)
Selfishness		-0.009		-0.324
		(0.015)		(0.531)
Low Incentive	-0.020	-0.024	-0.708	-0.834
	(0.043)	(0.043)	(1.475)	(1.470)
High Incentive	-0.025	-0.027	0.656	0.526
	(0.038)	(0.038)	(1.299)	(1.294)
Constant	0.496***	0.590***	7.831***	10.334***
	(0.041)	(0.051)	(1.415)	(1.772)
Observations	1,880	1,880	1,033	1,031
Cragg-Donald Wald F statistic	326.249	322.983	326.249	322.983

Notes: This table displays the coefficient estimates of IV regressions on the advisor's beliefs about the likelihood that Product B is of low quality, when there is a conflict of interest. The dependent variable (DV) in columns (1)-(2) is an indicator that takes value 1 if the advisor's belief that B is of low quality is directionally moving from 50 (the prior) towards the the Bayesian posterior. The DV in columns (3)-(4) is the distance between the advisor's belief and the prior of 50. This distance is positive if the advisor updates in the Bayesian direction, and negative otherwise. The advisor's preference to see the incentive first is instrumented by the random assignment to see the incentive first. Selfishness is the standardized number of choices of advisors in favor of a product for the client that gives them a commission, compared to an alternative that does not. This measure was elicited at the end of the experiment, using a multiple price list with 5 decisions. The regression models in columns (2) and (4) include individual controls for the advisor's gender and age. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

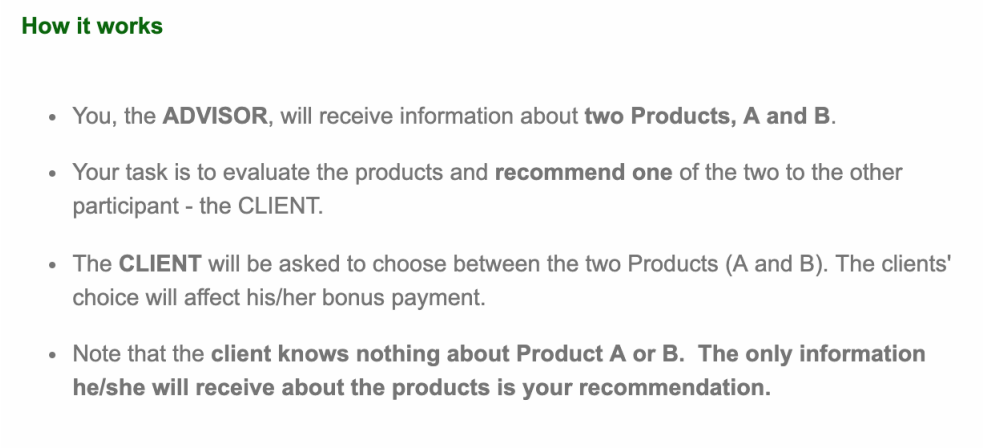
APPENDIX B: Instructions

The screenshots below present the instructions in the Choice Experiment, for the Incentives First Free treatment. The decision screen for the Predictions Experiment is shown as well. The instructions for all other treatments and experiments were based on this treatment, with the corresponding treatment modifications. Detailed instructions can be obtained from the authors.

Screen 1



Screen 2



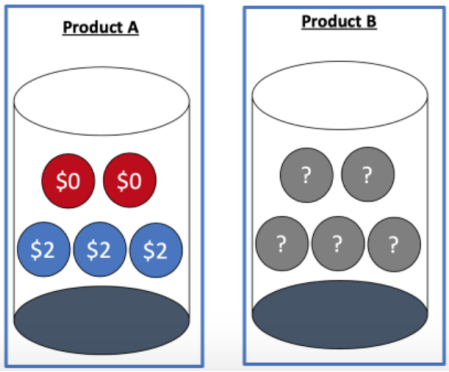
Screen 3 with attention questions

The Products

Product A and Product B are urns containing 5 payoff balls each. The payoff balls are either blue or red. Blue balls are worth $2 and red balls are worth $0. The combination of balls is different for the two Products, as described below:

- **Product A** is an urn with **3 blue ($2)** balls and **2 red ($0)** balls

- **Product B** is an urn that is either **high or low quality**, with equal chance.
  - If the Urn is **high quality** (50% chance), it has **four blue ($2) balls** (more than Product A).
  - If the Urn is **low quality** (50% chance), it has only **two blue ($2) balls** (fewer than Product A).
  - The quality of the urn was determined by the computer at random. **You do not know whether Product B is high or low quality for sure, but will soon receive information that will help you infer the quality of Product B.**



After receiving your recommendation, the client will choose one product between A and B. He/she will then randomly draw one ball from the urn. The payoff ball he/she draws will determine his/her payment.

Before you proceed, just a few questions to help you go over the instructions.

1. How much is a **red ball** worth?

- ○ $0.15
- ○ $1
- ○ $0
- ○ $2

2. How much is a **blue ball** worth?

- ○ $0.15
- ○ $1
- ○ $0
- ○ $2

3. How many **blue balls** are there in Product A?

- ○ 3 out of 5 balls are blue
- ○ 2 out of 5 balls are blue
- ○ 5 out of 5 balls are blue
- ○ 1 out of 5 balls is blue

4. The **quality of Product B** is **high** with...

- ○ 75% chance
- ○ 25% chance
- ○ 30% chance
- ○ 50% chance

5. Which of the following statements is correct? **Product B...**

- ○ ...is an urn with **4 blue balls ($2)** and **1 red ball ($0)** if its quality is **HIGH**, and it is an urn with **2 blue balls ($2)** and **3 red balls ($0)** if its quality is **LOW**
- ○ ...is an urn with **3 blue balls ($2)**and **2 red balls ($0)** if its quality is **HIGH**, and it is an urn with **3 blue balls ($2)** and **2 red balls ($0)** if its quality is **LOW**
- ○ ...is an urn with **5 blue balls ($2)** and **0 red balls ($0)** if its quality is **HIGH**, and it is an urn with **0 blue balls ($2)** and **5 red balls ($0)** if its quality is **LOW**

Before you proceed, **make sure you read these instructions carefully**. On the next screen, there will be one more question to verify that you paid attention. If you don't answer that question correctly, you will not be eligible to receive a bonus for this study.

Screen 4 with additional attention question

A Question for You

Before proceeding with your task, please answer the question below.

Imagine the client chooses **Product A**. What is the **chance** he/she gets **$2** (a **blue** ball)?

- ○ 1 in 5, because 1 out of 5 balls in Product A is **blue** ($2)
- ○ 2 in 5, because 2 out of 5 balls in Product A are **blue** ($2)
- ○ 3 in 5, because 3 out of 5 balls in Product A are **blue** ($2)
- ○ 5 in 5, because 5 out of 5 balls in Product A are **blue** ($2)

Screen 5

What You Know

- **You will soon receive more information that will help you gain some insights on the quality of Product B.**

- **The client does not know anything about Product A and B.** He/she will choose a Product after receiving your recommendation. The computer will then randomly draw a ball from the Product chosen by the advisor. The advisor will then will be paid accordingly.

Screen 6

Your payment

- Your task is to recommend either Product A or B to the client.
- You will receive **$0.50 for completing this study** and providing your recommendation.

- You may receive an **additional $0.15 commission depending on which product** you recommend.

- The **$0.15 commission** can be for recommending Product A **or** B. This has been determined randomly by the computer

Screen 7

Your choice

- You can choose to learn about your commission (i.e., whether product A or B yields a $0.15 commission) **before** or **after** obtaining information that will help you infer the quality of Product B. This information will be a ball randomly drawn from Product B. This ball will be placed back into the Urn.

- That is, you will choose between 2 options:
  - I want to learn which product recommendation gives me a **$0.15 commission (Product A or B) before** I obtain information that helps me infer the quality of Product B

    OR

  - I want to learn which product recommendation gives me a **$0.15 commission (Product A or B) after** I obtain information that helps me infer the quality of Product B

- Your preferred option will be implemented with 75% chance.

Screen 8

Your Choice

- Recall that you will receive $0.50 for completing this study and providing your recommendation.
- You may receive an additional $0.15 **commission** depending on which product you recommend.

What do you prefer?

- ○ I want to learn which product recommendation gives me a **$0.15 commission (Product A or B) before** I obtain information that helps me infer the quality of Product B
- ○ I want to learn which product recommendation gives me a **$0.15 commission (Product A or B) after** I obtain information that helps me infer the quality of Product B

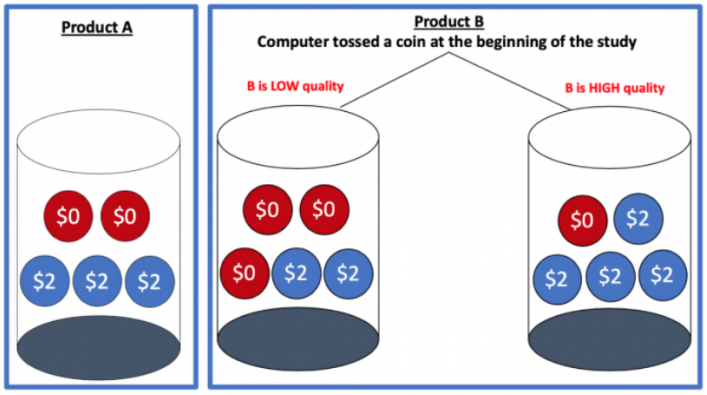
Screen 9 - Case of Random Assignment to Learn After

Following the procedure described above, you were assigned to learn about your commission **after** receiving information about Product B.

Screen 10

Next, you will obtain information that will help you infer the quality of Product B.

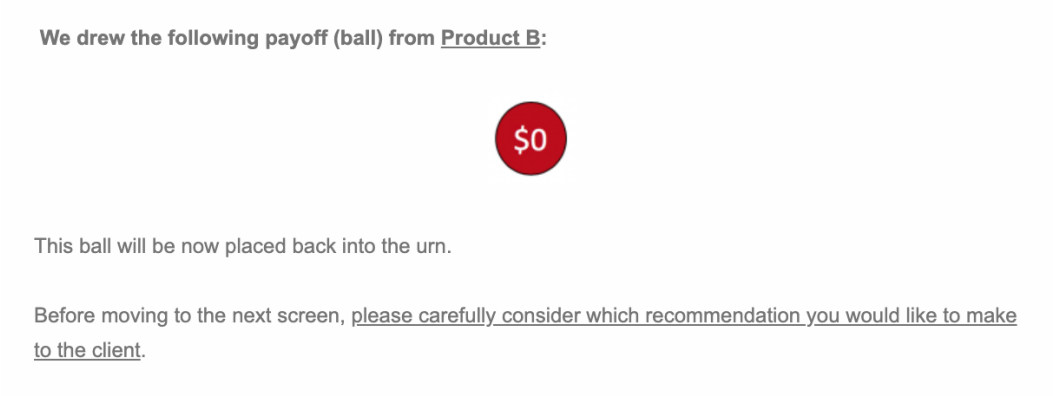
- **As a reminder, you know what Product A is. Instead, you don't know whether Product B is the low or high-quality Urn. Product B could have either High or Low quality with equal chance.** The combination of blue and red balls for both cases is depicted in the picture below. The quality of Product B was determined at random by the computer at the beginning of the study.



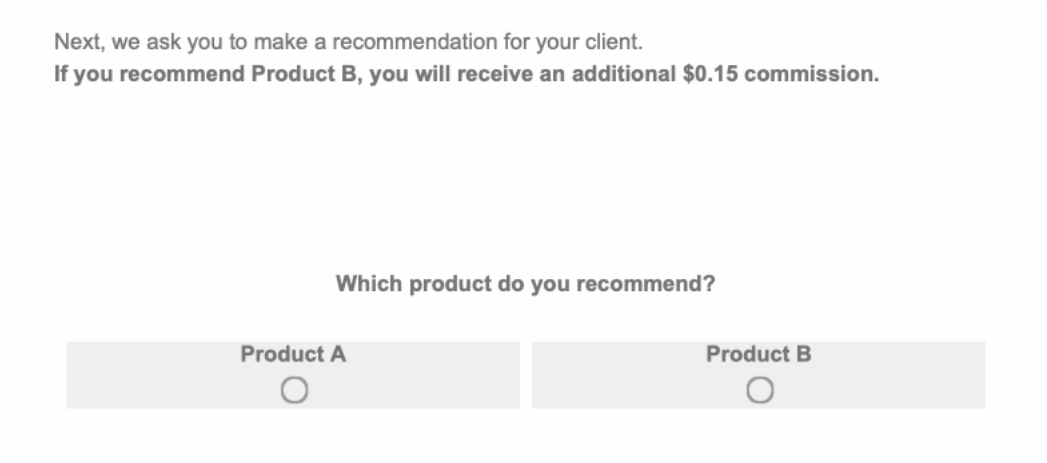
- **On the next screen, you will gain some insights on the quality of Product B. That is, we will randomly draw a payoff ball from Product B and display it on the next screen.**

- **After seeing the ball, you will be asked to choose which product, A or B, to recommend to the CLIENT.**

Screen 11



Screen 12 - Case in which Product B is incentivized



Prediction Experiment: Forecast Screen