

Cognitive Flexibility or Moral Commitment? Evidence of Anticipated Belief Distortion

Silvia Saccardo, Marta Serra-Garcia

Impressum:

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: <https://www.cesifo.org/en/wp>

Cognitive Flexibility or Moral Commitment? Evidence of Anticipated Belief Distortion

Abstract

Moral behavior is more prevalent when individuals cannot easily distort their beliefs self-servingly. Do individuals seek to limit or enable their ability to distort beliefs? How do these choices affect behavior? Experiments with over 8,900 participants, including financial and legal professionals, show preferences are heterogeneous - 30% of participants prefer to limit belief distortion, while over 40% prefer to enable it, even if costly. A random assignment mechanism reveals that being assigned to the preferred environment is necessary for curbing or enabling self-serving behavior. Third parties can anticipate these effects, suggesting some sophistication about the cognitive constraints to belief distortion.

JEL-Codes: D830, D910, C910.

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

Silvia Saccardo
Department of Social and Decision Sciences
Carnegie Mellon University
Pittsburgh / PA / USA
ssaccard@andrew.cmu.edu

Marta Serra-Garcia
Rady School of Management
University of California San Diego
San Diego / CA / USA
mserragarcia@ucsd.edu

May 2022

We would like to thank Johannes Abeler, Saurabh Bhargava, Christine Exley, Laura Gee, Russell Golman, David Huffman, Alex Imas, Michel Marechal, George Loewenstein, Kirby Nielsen, Theo Offerman, Ricardo Perez-Truglia, Peter Schwardmann, Jeroen van de Ven, Joel van der Weele, Lise Vesterlund, Eric van Damme, Roberto Weber, Florian Zimmerman, and participants at several conferences and workshops for helpful comments and suggestions. We would also like to thank Ehsan Amozegar, Daniel Henderson, Mandy Lanyon, Sarita Raghunath, Ben Schenk, Phillip Tan, and Stars Xu for excellent research assistance.

1 Introduction

The fundamental desire to preserve a positive identity often leads individuals to engage in motivated reasoning, distorting their beliefs to enable desired behaviors (e.g., Kunda, 1990; Bénabou and Tirole, 2006, 2011, 2016; Köszegi, 2006). The resulting belief distortion 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). Individuals can protect cherished beliefs by avoiding inconvenient information (e.g., Dana, Weber and Kuang, 2007; Golman et al., 2017). And, when information cannot be avoided, they can distort their beliefs ex-post through cognitive processes like attention and memory (e.g., Eil and Rao, 2011; Sharot, Korm and Dolan, 2011; Di Tella et al., 2015; Zimmerman, 2018; Huffman, Raymond, and Shvets, 2020; Pace and van der Weele, 2021; Möbius et al., 2022). Yet, there are limits to the ability to distort beliefs: belief distortion is enabled or constrained by contextual factors (Epley and Gilovich, 2016; Sloman, Fernbach, and Haggmayer, 2010). An important open question about motivated cognition is whether individuals anticipate these limits and, if so, whether they have preferences for enabling or constraining belief distortion. Do individuals attempt to limit belief distortion to commit to more accurate beliefs or would they rather seek out the cognitive flexibility needed to distort beliefs? How do these choices affect their subsequent behavior?

We investigate these questions in the domain of moral behavior (e.g., Abeler, Nosenzo and Raymond, 2019; Cohn et al., 2019), where there is evidence that individuals distort their beliefs to act self-servingly.¹ If informative signals cannot be avoided, belief distortion is enabled when individuals have “cognitive flexibility”: the cognitive ability to pay less attention to and therefore underweight potentially undesired signals. While previous findings suggest that some individuals may desire cognitive flexibility, little attention has been given to the possibility that some people may prefer to constrain belief distortion as a way to commit to moral behavior.

¹A 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, 2021), or by conveniently forgetting unpleasant news (Kouchaki and Gino, 2016; Saucet and Villeval, 2019; Carlson et al., 2020). These belief processes can lead to self-deception, enabling self-serving behavior (see, for example, Quattrone and Tversky, 1984; Bodner and Prelec, 2003; Bénabou and Tirole, 2006; Mijovic-Prelec and Prelec, 2010; Bénabou, Falk and Tirole, 2018).

In this paper, we investigate individuals’ willingness to constrain or seek out belief distortion, and study how being assigned to *experience* commitment or flexibility affects self-serving behavior.

We conduct a series of experiments where participants in the role of advisor ($N = 8,923$) 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 of information can constrain cognitive flexibility: Assessing what is best for another party without knowing one’s own incentives might raise attention to information about the other party’s outcome, committing individuals to a first unbiased judgment (e.g., Goldin and Rouse, 2000) and restricting the temptation to act self-servingly once new information is received (e.g., Babcock et al., 1995; Gneezy et al., 2020; Schwardmann, Tripodi and van der Weele, 2021). Consider experts – financial advisors, attorneys, accountants, expert witnesses, or reviewers – who have the ethical responsibility to make unbiased recommendations based on the information they receive but may succumb to the temptation of favoring their private interests. When evaluating new information (e.g., new investment funds, insurance policies, new cases or materials), experts who anticipate being tempted to violate their duty may actively commit to accurate beliefs by first assessing the information while being blind to their incentives. Or, they may actively seek out the cognitive flexibility needed to distort their beliefs by first examining potentially biasing information.

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 before making their recommendation but can choose the order with which they receive information. Seeing the quality signal first may increase the salience of this piece of information. It can draw attention to it and reduce a biased processing of the signal, thereby favoring self-serving recommendations.² To explain the effects of information order on behav-

²The 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), Köszegi and Szeidl (2013), 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). Some of this research has shown that new information not only

ior, we present a stylized theoretical framework that builds on Bénabou and Tirole (2002), in which quality signals receive more attention when they are seen first, in line with the literature on first impressions (Asch 1946, Anderson and Barrios, 1961; Anderson, 1965; Yates and Curly, 1986; Tetlock, 1983), work on anchoring and insufficient adjustment (e.g., Tversky and Kahneman 1974), and evidence on the effect of information order on self-serving behavior (e.g., Babcock et al., 1995). Advisors who first see the quality information pay more attention to quality signals and therefore have less scope to self-servingly suppress signals that are in conflict with the incentive, leading to less self-serving behavior. If ethical advisors anticipate that they may be tempted to provide a selfish recommendation, they may prefer to see the signal of quality first. By contrast, if they anticipate that they would like to enable self-serving information processing, they may prefer to see the incentive first and exploit the cognitive flexibility provided by this information order.

We begin by establishing that, in our context, exogenously assigning advisors to an information sequence affects their likelihood to engage in self-serving behavior. In line with prior work and the theoretical framework, when there is a conflict of interest, advisors are more likely to make recommendations that are in the client’s best interest when they assess the signal about quality first, compared to when they receive information about their incentives first. There is no effect of information order 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. First, we investigate preferences for information order. 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 (convenience) sample of online participants. Across both samples, we find substantial heterogeneity in preferences, which are split between wanting to see quality first and wanting to see the incentive first. If the choice is costless, 45% of advisors in the convenience sample and 55% of advisors in the sample of professionals commit to more accurate beliefs by choosing to see quality first (with the remaining 55% and 45%, respectively, seeking out cognitive flexibility). Since advisors’ preferences are close to 50%, a concern is that their preferences indicate indifference. However, indifference is not a prominent self-reported explanation of advisors’ choices of information order. Moreover, when we introduce costs, advisors reveal a strict preference: 30%

informs decision-making but it can also focus attention on certain beliefs.

of advisors are willing to incur a financial cost to receive quality information first, committing to more accurate beliefs, and 41% of advisors are willing to incur a financial cost to see the incentive first, pursuing cognitive flexibility.

Advisors' preferences to see quality information first are strongly correlated with advisors' morals, as measured in a separate task in which advisors always face a conflict of interest. Their preference for information order is also correlated with advisors' willingness to take up a stronger form of moral commitment: Advisors who prefer to assess quality first are more likely to blind themselves from learning about their incentive prior to their recommendation altogether. This evidence is in line with our theoretical framework and suggests that individuals anticipate that seeing quality first favors moral behavior.

Next, we investigate advisors' behavior: How does seeking out commitment or flexibility affect the rate of self-serving recommendations? To answer this question, in the experiment advisors' preferred information sequence is implemented with 75% chance. When advisors are assigned their preferred information sequence and are faced with a conflict of interest, there is a 19-20 percentage point gap in recommendations of the incentivized product between advisors who seek out flexibility and those who seek out commitment. However, there is no gap when advisors are not assigned to see information in their desired order. Conditional on preferences, being assigned to *experiencing* flexibility (vs. commitment) is crucial to advisors' ability to behave self-servingly, suggesting that behavior observed among those who are assigned their preferred information order does not just reflect sorting. For advisors who seek out commitment, we find that being assigned to see quality first significantly reduces self-serving recommendations. This result confirms that altering the order of information to assess quality first can be an effective moral commitment strategy.

The behavior of advisors who seek out flexibility speaks to an important open question in the philosophical discourse of self-deception: whether individuals can *intend* to self-deceive without rendering such intentions ineffective (Mele, 1987 and 2001; Bermúdez, 2000; see also Mijovic-Prelec and Prelec, 2010). A prominent hypothesis regarding the dynamics of self-deception is that actively seeking flexibility might prevent individuals from subsequently being successful at using this flexibility to self-deceive. If this is the case, seeking out flexibility by choosing to see information about one's incentive first may not work at enabling self-serving behavior. In contrast with this hypothesis, our results suggest that actively seeking flexibility

by choosing to see incentive information first does not impede advisors' ability to engage in self-serving behavior: advisors who prefer and are assigned to see the incentive first are significantly more likely to make the self-serving recommendation than those who seek out flexibility but are not assigned to experience it.

Advisors' beliefs about product quality are in line with their recommendations. Advisors who pursue and get cognitive flexibility exhibit beliefs closer to the prior, consistent with the assumption in our framework that signals of quality that are seen later receive less attention. We also find some evidence indicating that advisors display a directionally larger asymmetry in updating when seeing the incentive first. These findings are broadly consistent with advisors who seek and get cognitive flexibility being more able to dismiss informative signals, and in line with the theoretical predictions.

Advisors' preferences and recommendations are consistent with a proportion of them being sophisticated about the effect of information order on behavior. In two additional experiments, we provide evidence in support of this interpretation. First, we test whether preferences for cognitive flexibility or commitment respond to changes in advisors' incentives (see also, Coutts, 2019). Our data shows 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. Yet, when the gains from belief distortion further increase we do not see a similar increase in the demand of cognitive flexibility. This concavity is consistent with advisors experiencing less moral conflict as their incentives increase and, hence, the increase in demand for cognitive flexibility responding less to the incentive increase. Second, we test whether third party participants (the Information Architects, IAs) anticipate the effect of information order on advisors' behavior. In the experiment, IAs do not receive information but choose the order in which advisors learn about their incentives and the quality signal. We vary IAs' incentives to be aligned with the advisors' or the clients' payoffs, and ask them to choose the order of information for advisors. Our findings reveal that Information Architects are more likely to have advisors first assess quality without seeing the incentive when their own incentives are aligned to those of the client.

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). While

prior work has documented individuals' tendency to behave self-servingly despite an overall desire to feel moral (e.g., Gino, Norton and Weber, 2016), an open question is whether, in anticipation of the conditions that facilitate belief distortion, individuals desire to constrain belief distortion to uphold their morals. Our findings suggest that some advisors anticipate that changes to the way information is presented can constrain belief distortion, and that moral individuals are significantly more likely to take up opportunities for moral commitment, choosing to blind themselves from incentive information when making their initial judgments. At the same time, we find that a substantial fraction of individuals is willing to incur costs to seek out cognitive flexibility, when the gains from belief distortion are large enough.

Understanding how to mitigate the negative consequences of information asymmetries in presence of conflicts of interest (see, e.g., Darby and Karni, 1973; Crawford and Sobel, 1982; Pitchik and Schotter, 1987; Bénabou, 2013; Sobel, 2020; Malmendier and Schmidt, 2017), has important implications for the design of expert systems. Experts across a variety of professions – such as financial or legal professionals, expert witnesses, reviewers evaluating scientific research, and admission officers assessing candidates qualifications – are often called to make judgments that may be biased by private interests. A large literature has been concerned with designing decision-making environments in a way that prevent expert bias (e.g., Robertson and Kesselheim, 2016). We find that *temporarily* blinding experts from potentially biasing information can be a potential mechanism to reduce bias. Hiring managers and admission officers could be temporarily blinded from candidates information that is not relevant to their qualifications (e.g., Fath, Larrick and Soll, 2022) and review processes could be double-blind such that the identity of the authors of the research is initially unknown to the researcher evaluating them (Yankaeur, 1991). Financial advisors could also be informed about the commissions they receive when recommending a new investment product only after they have learned about the investment's characteristics. Even if, over time, such advisors may learn their incentives, first impressions can affect experts' quality assessments and thereby have a long-lasting effect on expert behavior (e.g., Chen and Gesche, 2017).

Our findings have implications for the self-selection of experts into organizations as well as for organizational design. We document substantial heterogeneity in preferences for commitment or flexibility, which could lead experts to self-select into types of organizations according to their practices or policies to prevent bias. Prior work has found evidence that social preferences correlate with selection into differ-

ent industries (e.g., the private or public sector: Gill et al., 2022; Serra, Serneels, and Barr, 2011; Hanna and Wang, 2017; Barfort et al. 2019). Even within the same industry, organizations differ in the way in which they design environments in which experts make decisions, and experts may choose to work in organizations whose policies align with their commitment or flexibility goals. Although such self-selection is important, our findings also suggest that self-serving behavior is more (or less) likely to arise when individuals effectively get to experience such flexibility (or commitment). This insight contributes to work recognizing how the immediate context where individuals make decisions can exert a strong influence in the extent of self-serving behavior, and urging organizations to design contexts where individuals are more likely to uphold their morals (Epley and Tannenbaum, 2017). Since those who make decisions in organizations often have some discretion in designing the informational structures and institutional arrangements that govern their behavior, our findings suggests that these individuals may make such design decisions with commitment or flexibility in mind.

2 Experimental Design

Our aim is to investigate individuals’ willingness to constrain or seek out belief distortion and examine how these choices affect self-serving behavior when unwelcome information cannot be avoided. Studying these questions requires an environment in which (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. Further, it requires an environment where (iii) individuals can actively pursue cognitive flexibility (or, conversely, mitigate it), when given the choice, 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.

2.1 The Advice Game

The advisor recommends one of two products, 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 \$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 quality 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 quality. 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 ($\iota = \$0.15$), 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 Main Experiments

We conduct four online experiments, as summarized in Table 1. We first present the two main experiments, NoChoice and Choice. In Section 3, we then present a stylized theoretical model that provides a lens through which to view the effect of information order in those experiments, guiding our main hypotheses. In Section 4, we describe the two additional experiments and further details of the experimental procedures.

A. The NoChoice Experiment. The goal of the first experiment is to establish that cognitive flexibility varies with the order of information. This experiment has two treatments. In the See Incentive First treatment, the advisor first receives

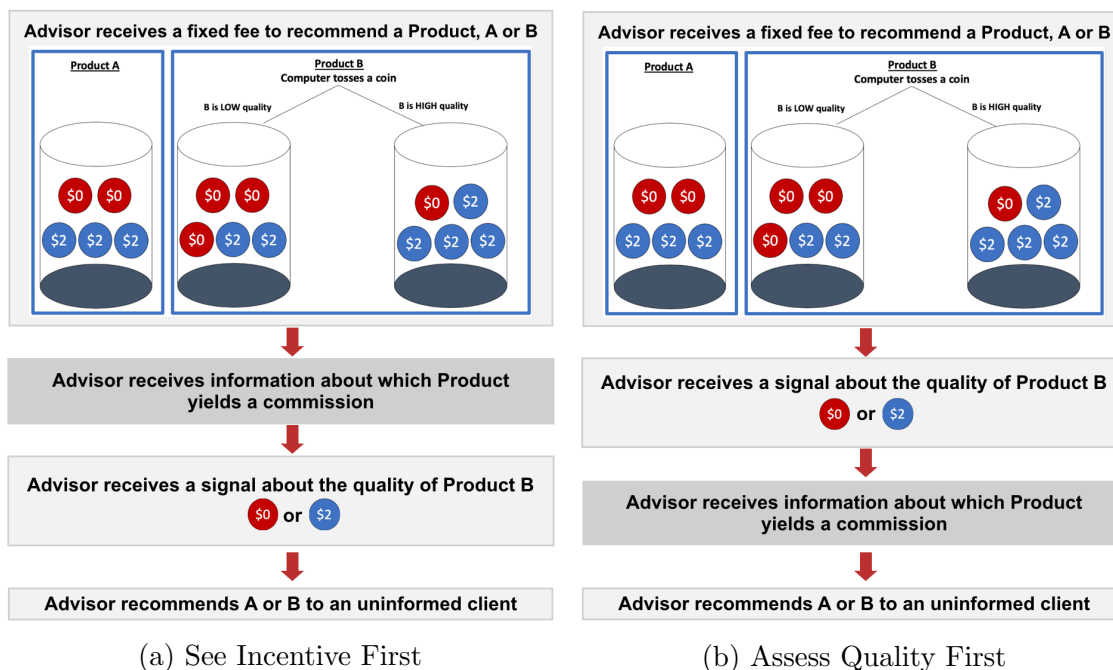


Figure 1: The Advice Game

information about which product recommendation is incentivized (Figure 1a) and then, on a later screen, the quality signal about product B. In the Assess Quality First treatment, the advisor first sees the quality signal about product B and only later, on the recommendation screen, learns about her incentive (Figure 1b). In both treatments, the evaluation of the signals only occurs in the advisors' mind. The incentive is always shown on the recommendation screen, with what varies being whether the incentive information also appears before the quality signal.³

B. The Choice Experiment. In this experiment we elicit advisors' preferences for information order in the advice game, and examine how being assigned to experience a given order affects recommendation decisions. To estimate the effect of information order on recommendations, conditional on advisors' preferences, advisors' choices are implemented probabilistically. With 50% chance, the advisors' choice is implemented, while with the remaining 50% chance, the advisor receives a 50/50 randomization. Advisors are informed that their preference is implemented with 75% probability. In this experiment, there are three conditions. In the *Choice Free* treatment, advisors make a simple choice between seeing the incentive first or assessing quality first. We conducted this experimental treatment with a sample of

³In Online Appendix D, we report the design and result of an additional wave of data collection that tests the effect of receiving information simultaneously.

Table 1: Experimental Design Outline

Experiment	Treatment	What do advisors see first?	N
Documenting Cognitive Flexibility: Information Order Affects Recommendations			
NoChoice	See Incentives First	Incentive	152
	Assess Quality First	Quality Signal	147
Preferences for Information Order: Cognitive Flexibility or Moral Commitment?			
Choice			
<i>Main Treatments</i>	Choice Free—Professionals	Advisor’s Choice	712
	Choice Free	Advisor’s Choice	2377
	Incentive First Costly	Advisor’s Choice	1358
	Quality First Costly	Advisors’ Choice	1067
<i>Robustness (in Appendix)</i>	Choice Free - High Stakes (10-fold)	Advisor’s Choice	275
	Choice Free - High Stakes (100-fold)	Advisor’s Choice	110
	Choice Free - Replication	Advisor’s Choice	385
	Choice Free - Deterministic	Advisor’s Choice	369
Additional Evidence			
ChoiceStakes	Low Incentive	Advisor’s Choice	484
	Intermediate Incentive	Advisor’s Choice	511
	High Incentive	Advisor’s Choice	478
Information-Architect	IA-Advisor	Third Party Choice	245
	IA-Client	Third Party Choice	253

Notes. This table summarizes the main experiments in the paper. Table B.1. in Online Appendix B provides the detailed description of the sample, waves, pre-registrations, and treatments in each of the experiments. The sample size in indicated in this tables refers to sample sizes after pre-registered exclusions.

individuals who self-report to work in industries in which advice is frequent—finance (including insurance), and legal services (*Choice Free—Professionals*) as well as with individuals from a 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.

To examine whether advisors have strict preferences to see the incentive first or to assess quality first, we introduce a cost of seeing the incentive first (*Incentive First Costly* treatment) and a cost for assessing quality first (*Quality First Costly* treatment), within the AMT sample. In each treatment, advisors forgo an additional payment, equivalent to a third of their commission (\$0.05), if they choose to see their incentive or the signal of quality first, respectively.

As part of this experiment, we conduct two robustness tests. First, we examine whether the probabilistic implementation of advisors’ preferences affects their recommendation behavior. We find that when implementing their preferences with

certainty (in the *Choice Free-Deterministic* treatment), the effect of information on recommendations is not significantly different from that observed for advisors who were assigned their preference (in the *Choice Free - Replication* treatment, which replicated the Choice Free treatment, see Online Appendix E). Second, a concern in the Choice experiment is that the incentives in the experiment are relatively small. Previous work has shown that even small incentives can influence expert decisions (Marechal and Thöni, 2019; DeJong et al. 2016; Malmendier and Schmidt, 2017) and that cognitive biases tend to persist across a variety of incentive sizes (Enke et al., 2021; see also Camerer and Hogarth, 1999). Since incentives for advisors and experts may vary in size and often be larger, we implemented two variations of the *Choice Free* experiment that increased the stakes in the experiment by a factor of 10 (*High Stakes - 10 fold*) or 100 (*High Stakes - 100 fold*). We find no significant change in the effect of information order on recommendations, suggesting that the results are robust to larger incentives (see Online Appendix C.4 for details).

3 Theoretical Framework

To explain how an advisor can leverage the order of information to restrict or enable self-serving behavior, we present a stylized theoretical framework. We adopt the framework of self-deception by Bénabou and Tirole (2002), based on attention management and an inner conflict in the advisor’s morality.⁴ To reduce notation, we modify the advice game to focus on the distinction between the presence or absence of conflict between the advisor’s incentive and the quality signal. In this simplified game, the signal the advisor can receive either indicates a conflict with the incentive ($\sigma = c$) or no conflict with the incentive ($\sigma = nc$). The prior likelihood that the signal is $\sigma = c$ is ϕ . We assume clients follow the advisor’s recommendation.

3.1 Limited and motivated attention

Attention is often limited (e.g., Kahneman, 1973) and motivated (e.g., Lang, Bradley, Cuthbert, 1997; Karlsson, Loewenstein and Seppi, 2009; Pace and van der Weele, 2021). The literature on first impressions indicates that it may be automatic to pay more attention to the first piece of information individuals receive (e.g., Asch 1946; Anderson and Barrios, 1961; Anderson, 1965; Yates and Curly, 1986; Tetlock,

⁴We thank the editor and review team for encouraging us develop a theoretical framework that formalizes our predictions and guides our analyses.

1983).⁵ We hence propose that cognitive flexibility varies with the order with which information is presented, because such order affects the likelihood with which the signal of quality is paid attention to or “encoded.” This assumption is in line with our empirical data, where advisors’ belief updating patterns are in line with the work on first impressions. Consistent with attention playing an important role, some advisors self-reported (in an open ended question) that seeing the incentive first “gives it more salience” or “might make me pay less attention to what I was learning” and that seeing quality first would make them “pay closer attention,” allowing them to “have better knowledge about the products” and preventing the incentives from “clouding their judgment.”

Seeing the signal of quality σ first ($f = q$) increases the likelihood that the advisor encodes (or remembers, pays attention to) this signal relative to seeing the incentive first ($f = i$). The reason is that, when the signal of quality is seen first, the incentive is not known, and the advisor is more likely to encode the quality signal. By contrast, seeing information about the incentive first leads the advisor to focus her attention on the incentive and pay less attention to the signal of quality. Formally, the probability that the quality signal is encoded $\lambda < 1$. Encoding of the quality signal is more likely when the quality signal is seen first $\lambda^q > \lambda^i$. If the signal of quality is encoded, it can be in conflict ($\sigma = c$) or not in conflict ($\sigma = nc$) with the incentive. If the signal is not encoded, the advisor does not know the signal, leading to $\sigma = \emptyset$. Incentive information is assumed to always be encoded, since all advisors are shown the incentive information on the recommendation screen.

3.2 Unstable morality

How much attention advisors pay to signals matters in the advice game, as it can enable or constrain self-serving recommendations by the advisor. Specifically, the advice game aims to capture the moral dilemma that arises when the option that the advisor is incentivized to recommend yields a lower expected payoff to the client (i.e., the immoral choice). Advisors who recommend the incentivized product may feel immoral, and experience a moral cost m . This moral cost can be viewed as akin to lying costs in sender-receiver games (e.g., Gneezy, 2005; Abeler, Nosenzo and Raymond, 2019), because a large majority of the clients follow advisors’ recom-

⁵Note that there is also a literature finding evidence of recency effects (Benjamin, 2019). Hogarth and Einhorn (1989) propose a belief-adjustment model for updating beliefs. Existing evidence in Gneezy et al. (2020) and in our first (NoChoice) experiment suggests that primacy effects dominate in the advice game we study.

mendations.

Many individuals care about behaving morally, but moral behavior is often unstable (for a review, see, Gino, Norton and Weber, 2016). Recent work highlights that acting self-servingly may be tempting for some individuals (e.g., Bénabou, Falk and Tirole, 2018), while others may fear being too generous. In the context of the advice game, individuals who feel conflicted about the right behavior may initially want to act selfishly or morally, but anticipate that once they learn about their incentive and the quality signal their recommendation may change (tempting them to act more morally or selfishly). In line with this intuition, our data show that when advisors are not assigned to see quality first, though they prefer it, they behave more self-servingly. Similarly, when advisors are not assigned to see the incentive first, though they prefer it, they behave more morally. Echoing this behavior several advisors report that the commission would tempt them to be less moral, e.g., *“I felt it was better to learn (my incentive) after so that I wasn’t tempted to make a decision out of greed.”*, while some advisors mentioned wanting to know the commission first to avoid feeling tempted to go with what was best for the client: *“I wanted to know which one had a commission upfront so I could be less tempted by the randomized drawing of product B.”*

To illustrate the advisor’s inner conflict, we adopt a dual-self framework (Bénabou and Tirole, 2002; Bodner and Prelec, 2003), by which Self 0 and Self 1 may differ in their moral costs. Specifically, moral costs are randomly drawn for Self 0 and Self 1. Let m_t be the moral cost of Self $t \in \{0, 1\}$. We assume that m_t is distributed uniformly on $[0, M]$, and independently drawn, with $M > \iota$, where ι is the advisor’s incentive payment. Self 0 manages attention to the signal of quality, knowing m_0 but not m_1 , while Self 1 makes the recommendation decision based on the signal received from Self 0. This stylized formulation of the inner conflict does not include an explicit concern for self-image (see, e.g., Bénabou, Falk and Tirole, 2018, which includes both self-image and temptation; and models of self-image by Bénabou and Tirole, 2011; and Grossman and van der Weele, 2017), to simplify exposition, while allowing Self 0 to worry that after the information is presented his moral preferences may change.⁶

At the beginning of the advice game, Self 0 encodes the signal of quality σ with probability λ^f and sends $\hat{\sigma}$ to Self 1. Based on $\hat{\sigma}$, Self 1 forms a belief about the

⁶Qualitatively similar predictions would result if Self 1’s moral costs would be modeled with a β (temptation) parameter relative to Self 0’s moral costs, e.g. $m_1 = \beta m_0$, where β could be larger or smaller than 1.

likelihood that the signal is in conflict with the incentive ($r(\hat{\sigma})$). Self 1 chooses whether to recommend the incentivized option ($x = 1$) and receive the incentive ι , or not ($x = 0$). Her utility is:

$$U_1(x|\hat{\sigma}, m_1) = [\iota - m_1 r(\hat{\sigma})]x.$$

From the perspective of Self 0, her utility at the recommendation stage may differ from that of Self 1 due to a difference in moral costs. Self 0 knows the signal that was encoded initially (σ), leading to:

$$U_0(x|\sigma, m_0) = [\iota - m_0 r(\sigma)]x.$$

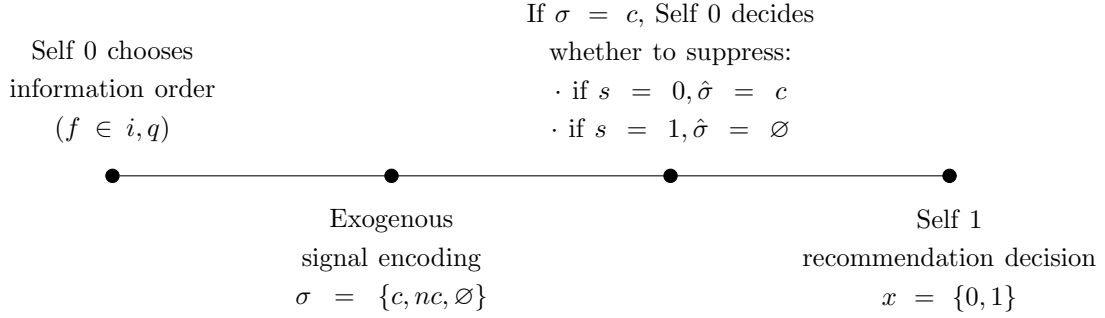
The potential conflict in moral costs between Self 0 and Self 1 may lead Self 0 to prefer to “manage” Self 1’s attention. If Self 0 starts the advice game with a high moral concern – that is, she initially draws high moral costs m_0 –, she may anticipate that her later Self 1 may have a lower moral concern (low m_1) and prefer “moral commitment.” Similarly, if Self 0 starts the advice game with a low moral concern (low m_0), this would motivate them to seek to pay less attention to informative signals about quality. This is easier when advisors have more “cognitive flexibility,” which we define as the ability of direct attention away from potentially undesired signals.

As in Bénabou and Tirole (2002), Self 0 can engage in motivated attention. When the quality signal σ is encoded and it is in conflict with the incentive, Self 0 chooses whether to “suppress” to the signal ($s = 1$) or not ($s = 0$). By suppressing, Self 0 attempts to act as if it was never encoded to begin with – a form of reality denial (see, e.g., Bénabou, 2015; Bénabou and Tirole, 2016). If Self 0’s signal is $\sigma = c$, Self 0 can choose $\hat{\sigma}$ to be c or \emptyset . Otherwise, $\hat{\sigma} = \sigma$. Suppressing an encoded signal is costless, and Self 0 suppresses with probability p_s . When Self 1 does not receive a signal (it is $\hat{\sigma} = \emptyset$), she uses Bayes’ Rule to form a belief about the likelihood that it is in conflict with the incentive, $r(\emptyset)$, as follows:

$$r(\emptyset) = \Pr(\sigma = c|\hat{\sigma} = \emptyset, f, p_s) = \frac{\lambda^f p_s \phi + (1 - \lambda^f) \phi}{\lambda^f p_s \phi + (1 - \lambda^f)},$$

where ϕ is the prior likelihood that the signal is in conflict with the incentive. If the signal received by Self 1 is in conflict with the incentive, then $r(c) = 1$, and if it is not in conflict with the incentive, $r(nc) = 0$. Figure 2 presents a timeline of the

model when Self 0 chooses the information order.



Notes: This figure shows, from left to right, the steps in the model when Self 0 decides the information order. Signal encoding occurs exogenously, depending on the information order f . If $\sigma = c$, Self 0 decides whether to suppress the encoded signal. Last, Self 1 makes her recommendation decision. If Self 0 does not decide (NoChoice experiment), the first step would be removed.

Figure 2: Timeline of the model, when advisors choose the information order

If Self 1 receives a signal that is in conflict with the incentive $\hat{\sigma} = c$, Self 1 chooses $x = 1$ only if $m_1 \leq \iota$. If the signal received is not in conflict with the incentive, there are no moral costs and Self 1 chooses $x = 1$. If Self 1 does not receive a signal, her inference about $r(\emptyset)$, the risk of recommending a product that is in conflict with the incentive, determines her decision to recommend the incentivized product. She recommends the incentivized product if:

$$m_1 \leq \frac{\iota}{r(\emptyset)}.$$

As in the experiment, we assume that, if Self 1's belief about the likelihood that the signal is in conflict with the incentive is the same as the prior, Self 1 recommends the incentivized product, i.e., $\iota - \phi M > 0$.

3.3. No Choice of Information Order

We start by considering first the case where Self 0 cannot choose the information order (as in the NoChoice experiment). Self 0 is assigned to see the incentive first or the signal of quality first. When Self 0 has a high moral cost, $m_0 > \iota$, she always conveys the signals that are encoded and never suppresses. This minimizes the likelihood that Self 1 recommends the incentivized product when the signal is in conflict with the incentive, providing a form of “moral commitment” by increasing attention to the quality signal.

When Self 0 is selfish and has low moral costs, $m_0 < \iota$, Self 0 has an incentive

to suppress signals that are in conflict with the incentive. Denote the probability that Self 1 recommends the incentivized product when she does not receive a signal of quality ($\hat{\sigma} = \emptyset$) by $q = \frac{\iota}{r(\emptyset)M}$. The expected utility of Self 0 from choosing to suppress with p_s is

$$E(U_0) = \lambda^f \left((1 - \phi)\iota + \phi((1 - p_s)((\iota - m_0)\frac{\iota}{M}) + p_s(\iota - m_0)q) \right) + (1 - \lambda^f)(\iota - \phi m_0)q.$$

Self 0 suppresses the signal of quality as often as possible, as long as Self 1 still recommends the incentivized product, and hence chooses,

$$p_s^* = \min\left\{\frac{(1 - \lambda^f)(\iota - \phi M)}{\lambda^f \phi (M - \iota)}, 1\right\}.$$

Because the signal of quality is encoded less often when the incentive is seen first ($f = i$), this information order provides more “cognitive flexibility.” Self 0 can exploit the lower attention to engage in more motivated attention (suppression). This result is summarized in Proposition 1 (further details in Online Appendix A).

Proposition 1. *When the signal of quality is shown first, the advisor is less likely to suppress it and less likely to recommend the incentivized product when it conflicts with the incentive, than when the information about the incentive is shown first.*

Hence, when there is a conflict of interest, the likelihood of recommending the incentivized product increases when advisors see their incentive first. When there is no conflict of interest, the advisor recommends the incentivized product under both information orders.⁷ This yields our first hypothesis.

Hypothesis 1 (NoChoice experiment). If advisors are assigned to See the Incentive First, the likelihood with which advisors recommend the incentivized product when the signal is in conflict with the incentive is higher than when they are assigned to See Quality First.

⁷This result highlights that the difference between recommendations is expected to be present when the signal is in conflict with the incentive, due to our focus on the role of attention management to signals as the mechanism through which information order affects recommendations. A difference in recommendations when there is no conflict of interest may also arise if advisors who see the quality signal first are less likely to pay attention to incentive information. We find little evidence for this in our data. If there is no conflict of interest, the difference between information orders is either absent or small, between 20 to 30 percent of that observed when there is a conflict of interest.

3.3 Advisor’s Choice of Information Order

Given the effects of information order on the attention process and recommendation decisions, what order of information does the advisor prefer?⁸ We first consider the case of a sophisticated advisor who correctly anticipates the decrease in attention when the incentive is seen first. As shown in Proposition 2, if Self 0 has low moral costs, she prefers to see the incentive first, since it affords more “cognitive flexibility”. In contrast, if Self 0 has high moral costs, she prefers to see the quality signal first to have more “moral commitment.”

Proposition 2 (Sophisticated Advisors).

- *If Self 0 is selfish ($m_0 \leq \iota$), she chooses to see the incentive first ($f^* = i$). This order increases the likelihood that Self 1 recommends the incentivized product when the signal is in conflict with the incentive.*
- *If Self 0 is moral ($m_0 > \iota$), she chooses to see quality first ($f^* = q$), which decreases the likelihood that Self 1 recommends the incentivized product when the signal is in conflict with the incentive.*

How would this prediction change if advisors are not sophisticated about the malleability of attention? We define a naïve advisor as one who believes that the order of information does not affect attention. Formally, the advisor believes that her attention is as limited when seeing the incentive first as when seeing quality first, $\hat{\lambda}^q = \hat{\lambda}^i = \lambda^i < 1$. If the advisor were naïve, then she would not anticipate any effect of information order, leading to Proposition 3.

Proposition 3 (Naïve Advisors). *If the advisor does not anticipate the effect of information order on attention, she is indifferent between seeing the incentive first or seeing quality first.*

These results yield Hypothesis 2, for the Choice experiment.

Hypothesis 2 (Choice experiment).

- (a) Preferences: If advisors are sophisticated, those who are more selfish (lower moral costs) are willing to pay to see the incentive first, while advisors who are more moral (higher moral costs) are willing to pay to see quality first. If advisors are naïve, they are not willing to pay for any information order.

⁸We assume that the advisor’s choice is implemented with certainty to simplify exposition.

- (b) Recommendations: Advisors who actively choose (and pay) to see the incentive first are more likely to recommend the incentivized option if the signal is in conflict with the incentive. Conversely, advisors who choose (and pay) to assess quality first are less likely to recommend the incentivized option if the signal is in conflict with the incentive.

The theoretical framework we proposed relies on two simplifying assumptions the validity of which we explore in the data analyses. The framework assumes that advisors' active choice of information order does not restrict their ability to suppress signals that are in conflict, since suppression depends on information order alone and not on whether the information order was chosen or assigned to the advisor. Philosophers, however, have argued that intentionality can decrease the scope for self-deception (e.g., Mele, 1987 and 2001).

Our experiments allow us to better understand the role of intentionality in belief distortion, which we test in two ways. First, we examine whether advisors who prefer to see the incentive first and are assigned their preferred order recommend the incentivized product more often than advisors who prefer to see the incentive first and are not assigned their preferred order. This comparison allows us to test whether advisors who choose to see the incentive first are still able to distort recommendations when they *experience* more cognitive flexibility, although they potentially intended to self-deceive. Second, focusing on advisors who are assigned their preferred order, we test whether the gap in recommendations between advisors who prefer to see the incentive first and those who prefer to see quality first is larger in the Choice experiment than in the NoChoice experiment. Since advisors select according to their preference in the Choice experiment, but they do not in the NoChoice experiment, comparing the difference in recommendations between the experiments allows to measure whether information order affects advisors similarly when such information order is directly chosen by advisors.

The theoretical framework also assumes belief distortion occurs through advisors' limited and motivated attention to the signal of quality of product B. Belief distortion in the advice game, however, need not (only) be about quality. Research on self-serving biases has also shown that individuals may distort their beliefs about what is fair in a self-serving manner (e.g., Babcock et al., 1995; Gneezy et al., 2020) allowing them to maintain a self-image as moral (e.g., Bénabou and Tirole, 2011; Grossman and van der Weele, 2017; Bénabou, Falk and Tirole, 2018). Our framework complements these accounts by focusing on belief distortion that occurs

through attention to quality signals. If belief distortion takes place through attention to quality signals, we would expect that advisors who choose to see the incentive first hold beliefs regarding the quality of the product that are closer to the prior (as they pay less attention) than those of advisors who prefer to assess quality first. Since lower attention allows more suppression of signals that are in conflict with the incentive, advisors who choose to see their incentive first should then be more likely to exhibit directional bias in updating.

4 Additional Experiments and Procedures

4.1 Additional Evidence of Anticipation

Advisors' choices in the Choice experiment could be driven by a variety of motives. We conduct two additional experiments to provide complementary evidence that choices of information order respond to incentives and are consistent with individuals anticipating the effect of pursuing (and getting) cognitive flexibility or commitment.

A. The ChoiceStakes Experiment. We test whether advisors' preference to see the incentive first respond to their incentive to recommend the incentivized product. If the gains from flexibility decrease (via a substantial decrease in the advisor's incentive), and advisors are sophisticated, we would expect their preference to see the incentive first to drop. In contrast, if the advisor's incentive increases, the effect of her preference is ex-ante unclear. Their preference to see the incentive could increase, since there is a larger gain from flexibility. But, their preference could be weakened, if the incentive is large enough, such that the advisor does not perceive the advice decision as presenting a moral dilemma (see Online Appendix A for details). In the experiment, we keep the payoffs for the client the same, and vary the incentive 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 double it 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.

B. The Information Architect Experiment. We introduce third-party participants in the role of Information Architects (IAs), who are matched with an advisor and choose the order in which advisors receive information in the advice game. In these treatments, IAs see the same general instructions seen by the advisors and are

informed that they will decide how an advisor receives information in the experiment (see Instructions in Online Appendix G). To investigate whether IAs anticipate the effect of information order on behavior, we vary the IAs’ incentive to have the advisor recommend one of the two products and test whether aligning the IAs’ incentive with those of the advisor (as opposed to those of the client) affects their choices of information order. In the *IA-Advisor* treatment, IAs’ incentives are aligned with those of the advisor, receiving a \$0.15 payment if the advisor recommends the incentivized product. In the *IA-Client* treatment, IAs’ incentives are aligned with those of the client, receiving a \$0.15 payment if the advisor recommends the product with the highest expected payoff for the client. We examine whether IAs’ choice of information order for the advisor varies with these incentives, suggesting that they anticipate the effect of information order on the advisor’s behavior. In this experiment, only the advisor receives the information (i.e., the signal of quality and the information about what product yields a commission); the IA chooses the order of information for the advisor without receiving such information herself. In doing so, we can remove curiosity from driving preferences for information order. To further examine whether individuals anticipate the effect of information order on recommendations, in Online Appendix F we report an additional experiment where we ask third party individuals to predict recommendation rates under the different information orders (see, e.g., DellaVigna and Pope, 2018; DellaVigna, Pope, and Vivaldi, 2019).

4.2 Experimental Procedures

We conducted all experiments except the Choice Free—Professionals treatment, on Amazon Mechanical Turk (AMT).⁹ All Experiments on AMT were pre-registered on aspredicted.org. The Choice experiment was conducted in three waves, with each wave pre-registered separately. Online Appendix B provides pre-registration numbers, detailed design information, recruitment procedures, and exclusion criteria for the experiments. The sample of professionals was drawn from individuals who self-report to work in two industries in which advice is very frequent: finance and insurance, and legal services. We used Prolific Academic (Palan and Schitter,

⁹Existing 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). AMT allows us to recruit a large sample of participants and explore the mechanisms behind the choices advisors make in additional experiments.

2018) and CloudResearch (Litman, Robinson, and Abberbock, 2016) to target the experiment to professionals in these industries.¹⁰

All experiments were conducted online, using Qualtrics surveys. Participants received a base payment of either \$0.50 or \$1 for participating in a 5-7 minute study. As detailed above, in most of the advice game experiments, all advisors received a \$0.15 commission depending on their recommendation, and one out of 10 advisors was matched with a client. In the Choice Free—Professionals treatment and in a group of participants in the Choice Free AMT treatment, we kept the expected value for the advisor the same, but implemented a probabilistic payment structure. We paid 1 out of 100 advisors a \$15 commission, and matched all of the randomly selected advisors with a client. In these treatments, the payoffs of product A or product B were scaled up to \$0 or \$20.¹¹

Since our main interest is in the cases where advisors faced a conflict of interest, we predetermined which product yielded a commission in a way that maximized the number of cases in which advisors faced a conflict of interest. All advisors randomly assigned to having a low-quality product B received a commission for recommending product B; 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. 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.

At the end of the experiments, we randomly selected advisors according to the

¹⁰Prolific 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 ($p=0.308$), and recommendations did not differ either ($p=0.820$). Concurrent work focusing on truth-telling among financial professionals on Prolific and a proprietary pool consisting of financial professionals (portfolio managers, financial advisors, etc.) found similar behavior across pools (Huber and Huber, 2020). Professionals self-reported their job titles in 95% of the cases (677 of 712). Two independent raters examined whether their job titles involved a fiduciary duty or not, and found that a majority of professionals' jobs (61.9%) were considered to have fiduciary duty. See Online Appendix B.2 for further details.

¹¹To test whether advisors display different responses to probabilistic incentives, in one of the waves of the Choice experiment, we recruited 1,053 participants and randomized whether incentives were probabilistic as in the professional sample and whether the incentivized product was presented on the left side or the right side of the screen. We found no effect of incentive size, order or their interaction on the preference to see the incentive first (t -stat = -1.46 , $p = 0.144$ for incentive size, t -stat = 1.41 , $p = 0.159$ for order, and t -stat = -0.03 , $p = 0.980$ for the interaction of the two). We also found no effect of incentive size, order or their interaction on recommendations (t -stat = 0.34 , $p = 0.733$ for incentive size, t -stat = 0.45 , $p = 0.652$ and t -stat = 0.85 , $p = 0.396$ for their interaction). Hence, we pool the data and control for these design variations in all regression analyses.

procedures of each experiment and sent each advisor’s recommendation to a client. We recruited clients ($N = 924$) later and informed them that advisors had received information about the two products and had made a recommendation.¹² Clients saw their advisor’s recommendation and then made a choice between the two products; they received no other information about the products. Overall, 84% of clients followed the advisor’s recommendation.

4.2.1 Additional measures

After the recommendation stage, we collected additional measures.

Beliefs. We elicited advisors’ beliefs about the likelihood that the quality of B was low by asking advisors i) to choose one of ten 10 percentage-point intervals, 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.¹³

Moral costs. We measured advisors’ moral concern for providing a recommendation that helps the client, when there is a conflict of interest, using a multiple price list, in all experiments except for *Choice-Free Professionals*. Advisors made five recommendation decisions to a newly matched participant, the “advisee.” There were two products, X and Y. Product Y had the same payoffs as product B in the experiment. Advisors were incentivized to recommend Y, with a \$0.15 commission, and received a signal of quality of product Y that indicated that a \$0 had been drawn from 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. Given the payoffs of X, recommending Y harmed the client if X paid \$2 with a probability of 0.6 or higher. In those decisions, if the advisor chose to recommend Y, she could suffer moral costs. We consider this measure captures the moral costs of Self 0, within the theoretical framework, because the signal of quality was presented at the same time as products X and Y. For simplicity, we refer to this (standardized) measure as the advisor’s overall selfishness.¹⁴ In all of the main analyses, following our pre-registrations, we

¹²Following the instructions, we recruited 1 out of 10 clients for all treatments other than the Choice Free Professionals treatment and a subsample of the Choice Free treatment in the second wave of the experiment, where we recruited 1 out of 100 clients, and the High Stakes - 100 fold treatment where we matched each advisor with one client.

¹³The payment was \$15 in the Choice Free treatments in which 1 out of 100 advisors was selected for payment

¹⁴At the end of the experiment, we randomly selected one out of 10 advisors, randomly picked one of the 5 recommendations, and showed them to a client. For this purpose, we recruited a total

focus on attentive advisors who gave consistent responses in this task, excluding those who switched multiple times. The results remain qualitatively similar if we include them, as shown in Online Appendix C.

Blinding. In the third wave of data collection of the Choice Experiment, we measure take up of a stronger form of moral commitment in a separate task. In this task participants are assigned to the role of advisor and can choose to blind themselves to the incentive information – i.e., not know which product they are incentivized to recommend — prior to providing a recommendation to another participant in the role of advisee.¹⁵ Advisors knew that the incentive and the signal of quality would be drawn again. Advisors could choose to blind themselves and receive information about their incentive only *after* providing her recommendation, or they could choose not to blind, which implied that they received information about the signal of quality and the incentive at the same time, before providing her recommendation. We consider preferences for blinding in this task as a stronger form of moral commitment than choosing to see quality first because advisors choose not to know their incentive at all prior to their recommendation decision.

Explanations of Choice. In the second wave of data collection of the Choice Experiment (Choice Free treatment) and among *Choice Free–Professionals*, we added an open ended question asking participants to explain how they made their decision about order of information. Two independent raters, blind to advisors’ choices, coded the responses of 1,747 advisors (including $N = 712$ professionals) and classified their responses into four categories, which apply to 91% of the responses. The remaining 9% consists of empty or unrelated comments according to both raters. The four categories were “limiting bias” (i.e., messages that explicitly stated that the reason for their preference was to be less biased in the evaluation, a measure of preference for commitment to accurate beliefs and behavior), “indifference” (i.e., messages in which advisors stated that information order did not matter), “commission,” (i.e., messages in which advisors who indicated explicitly that they cared only about their own commission) or “other reasons,” (i.e., which captured whether advisors indicated that gut feeling, wanting information sooner, or other reasons

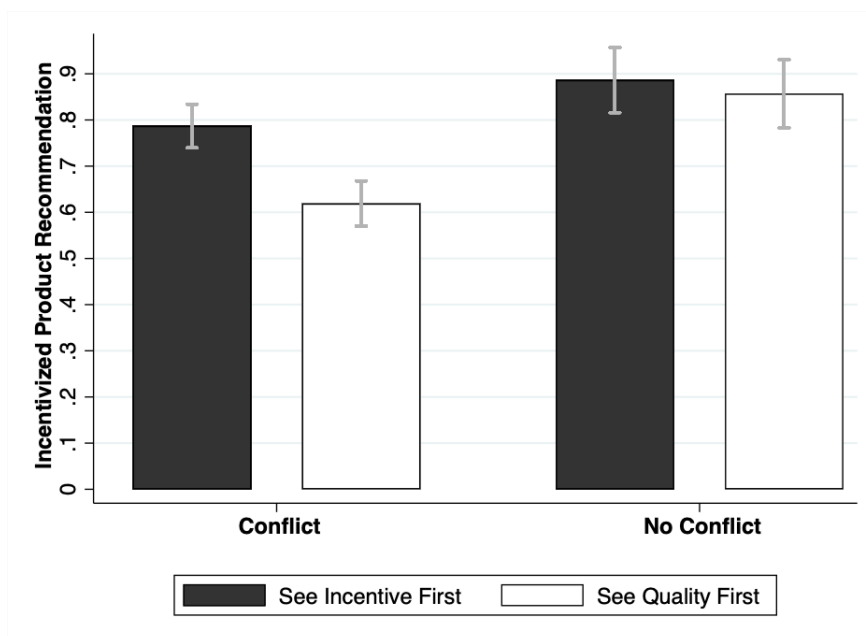
of 866 clients across all the experiments reported in Table 1. Of these, 80% of clients followed the advisor’s recommendation.

¹⁵At the end of the experiment, we randomly selected one out of 10 advisors and send their recommendation to an advisee. For this purpose, we recruited 188 advisees. Of these, 84% followed the advisors’ recommendation.

guided their preference). We did not expect advisors to openly express wanting cognitive flexibility in their comments. Consistent with this, such comments were rare in the data. We allowed coders to indicate multiple categories, though this was rarely done (in less than 3% of the cases). The two independent raters (see Online Appendix B for the coding procedures) 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.

5 Does the Order of Information Affect Advice?

We first test whether exogenously assigning a given order of information affects advice in the NoChoice experiment.



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, $N = 213$ for cases of conflict and $N = 86$ for cases of no conflict.

Figure 3: Recommendation of Incentivized Product, by Treatment

When advisors face a conflict of interest (i.e., the quality signal is in conflict with their own incentive), the rate of self-serving recommendation depends on the order in which information is presented to them. Figure 3 shows that in the See Incentive

First treatment, 79% of advisors recommend the incentivized product. In the Assess Quality First treatment, 62% of advisors recommend the incentivized product. This 17 percentage point difference is significant (Z -stat = 2.69, $p = 0.007$, $N = 213$). 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 89% of the time, while those in the Assess Quality First treatment recommend the incentivized product 86% of the time (Z -stat = -0.41 , $p = 0.685$, $N = 86$). These results are robust to controlling for demographics and for controlling for advisor’s selfishness (see Online Appendix C.1).

Throughout, advisors exhibit a preference for product A, recommending it 16% of the time, even when the quality signal is a \$2 and the advisor is incentivized to recommend B. Nevertheless, the effect of information order on behavior is similar regardless of what product is incentivized.¹⁶

Consistent with Hypothesis 1, these results suggest that, when there is a conflict of interest, seeing the incentive first provides more cognitive flexibility, enabling advisors to recommend the incentivized product more often than when the signal of quality is assessed first.

Result 1. When there is a conflict of interest, advisors who are assigned to See the Incentive First are significantly more likely to recommend the incentivized product than advisors who are assigned to See Quality First.

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

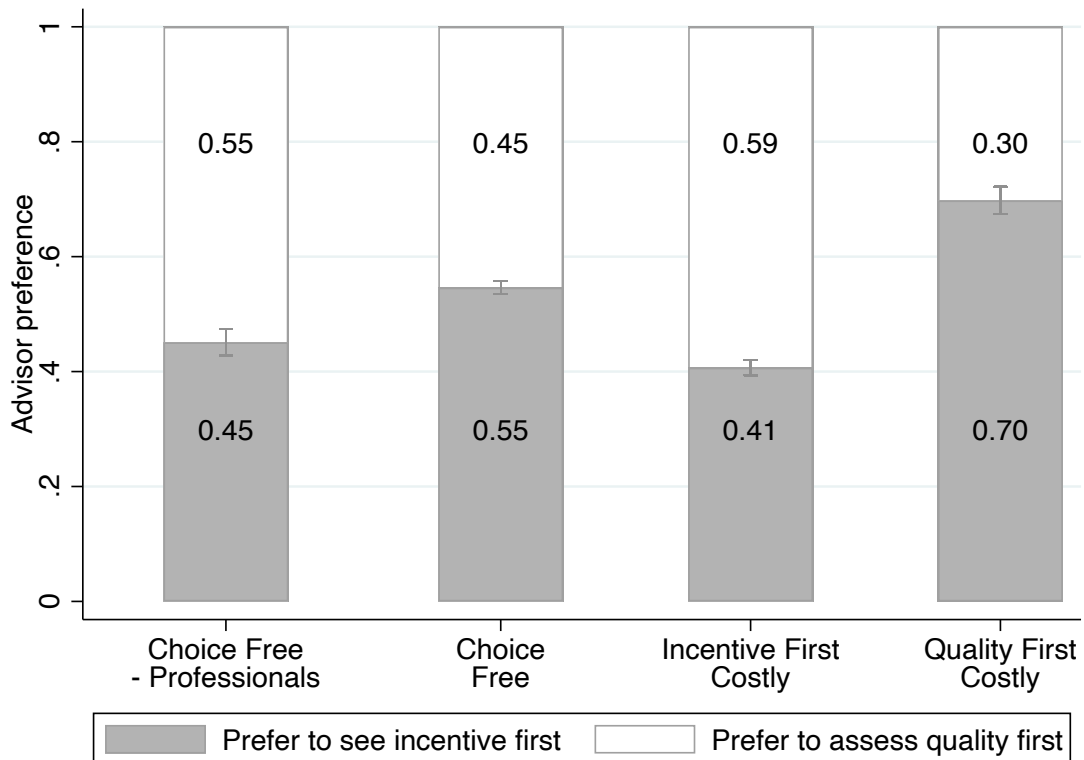
6 Preferences for Information Order: Cognitive Flexibility or Moral Commitment?

When choosing the information order is free, advisor preferences for information order are split between seeing the incentive first and seeing quality first, as shown

¹⁶In Online Appendix D, we report data from the additional wave of the study that tests effect of presenting both information about incentives and the quality signal simultaneously. The results show that, when both pieces of information are presented on the same screen, advisors behave similarly to the *See Incentive First* treatment, suggesting that in order for advice to be less influenced by incentives, advisors need to first process the quality signal without knowing what their incentives are.

in Figure 4. Since we conducted the experiment in several waves, the figure shows covariate-adjusted demand, controlling for wave, advisor gender and age (disaggregated results are shown in Online Appendix C.2).

Among professionals, 45% of advisors prefer to see the incentive information first, and among AMT participants, 55% of advisors exhibit the same preference. Conversely, between 55% and 45% of advisors choose to see the quality signal first, indicating that a substantial fraction of advisors would rather delay information about their own incentive.



Notes: This figure presents covariate-adjusted demand of advisors to see the incentive first or assess quality first estimated using OLS-regression. The covariates include the wave of data collection in the Choice experiment, whether the incentives were probabilistic or not, whether product *A* was presented on the left-hand side of the screen, the age and the gender of the advisor. Preferences by experimental wave are shown in Online Appendix C.2. Error bars indicate ± 1 SE.

Figure 4: Advisor Preference

When seeing the incentive first is costly, 41% of advisors 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 of advisors shows a strict preference. Similarly, when see the quality signal is costly, 30% of advisors are willing to pay a cost to see the quality signal first, which we interpret as a form of moral commitment to accurate beliefs. Compared to when choice is free, when seeing the incentive first is costly, there is a 14-percentage-point drop in demand to see the incentive first ($t - \text{stat} = -7.84, p < 0.001$), as shown in Table 2. When seeing quality first is costly, there is a 15-percentage-point increase in demand to see incentive first ($t - \text{stat} = 5.17, p < 0.001$).

Table 2: Preference for Information Order

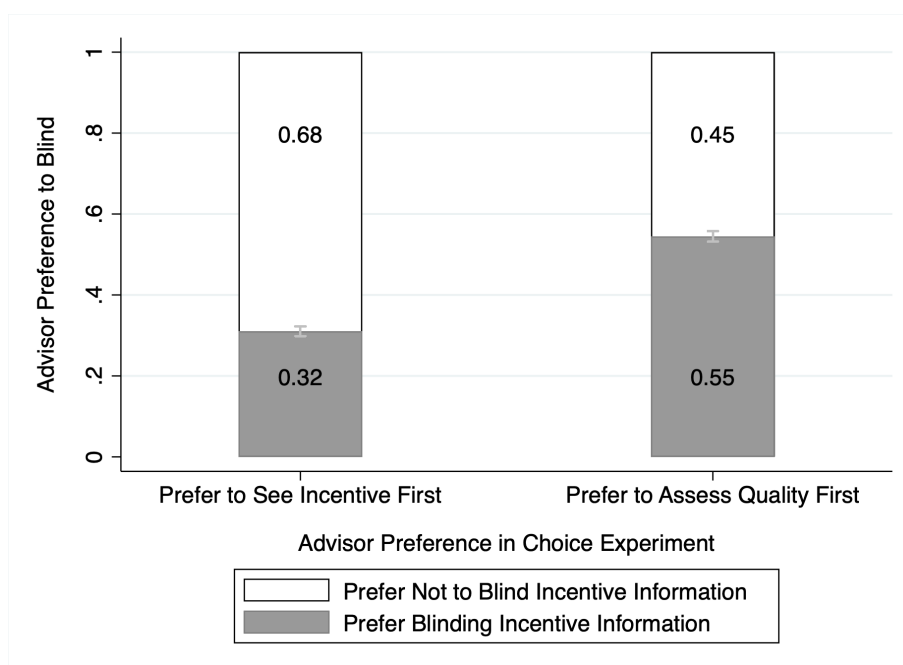
	(1)	(2)	(3)
	Prefer to See Incentive First		
See Incentive First Costly	-0.139*** (0.018)	-0.140*** (0.018)	-0.140*** (0.018)
Assess Quality First Costly	0.152*** (0.029)	0.152*** (0.029)	0.152*** (0.029)
Choice Free – Professionals	-0.095*** (0.026)		
Selfishness		0.028*** (0.007)	0.039*** (0.009)
See Incentive First Costly X Selfishness			-0.022 (0.016)
See Quality First Costly X Selfishness			-0.021 (0.018)
Female	-0.029** (0.013)	-0.024* (0.014)	-0.023* (0.014)
Age	-0.003*** (0.001)	-0.002*** (0.001)	-0.002*** (0.001)
Constant	0.674*** (0.024)	0.662*** (0.025)	0.661*** (0.025)
Observations	5908	5196	5196
R^2	0.034	0.040	0.040

Notes: This table displays the estimated coefficients from linear probability models on the preference to see the incentive first. See Incentive First Costly and Assess Quality First Costly are indicator variables that take value 1 in the respective treatment, 0 otherwise. Selfishness was elicited at the end of the experiment, using a multiple price list (MPL) with 5 decisions. The variable is a standardized measure of the number of times the advisor chose to recommend the incentivized product in the MPL task. The regression models in columns (2) and (3) include individual controls for the advisor’s gender and age, each wave of the experiment, whether incentives were probabilistic, the position of the products on the screen and the interaction between these two variables. Robust standard errors (HC3) in parentheses. * $p < .10$; ** $p < .05$; *** $p < .01$

Table 2 shows the determinants of the preference to see the incentive first, and columns (2)-(3) investigate its relationship with advisor selfishness. In line with Hypothesis 2(a), advisors who make more more selfish choices in the task designed to measure advisors’ moral costs prefer to see the incentive first significantly more often.

6.1 Preferences for Information Order and Preferences for Blinding

To examine whether advisors' preference to assess quality first is indicative of a desire for moral commitment, we test whether the preference to see quality first predict take up of a stronger form of moral commitment: choosing to blind oneself from incentives altogether. For this purpose, we focus on the subset of participants who took part in the blinding task. Advisors who prefer to assess quality first are significantly more likely to also prefer to blind themselves in the blinding task. As shown in Figure 5, 54.5% of advisors who choose to assess quality first also prefer to blind themselves. This fraction is significantly smaller, 31.0%, for advisors who prefer to see the incentive first (Z -stat = 9.11, $p < 0.001$, $N = 1484$).



Notes: This figure presents the fraction of advisors who chose to blind themselves from the incentive information, in the blinding task, and those who chose not to blinding themselves, conditional on their preference for information order in Wave 3 of the Choice experiment ($N=1484$). Error bars indicate ± 1 SE.

Figure 5: Take up of Blinding by Preferences for Information Order

The difference in preference to blind between advisors who prefer to see incentive first and those who prefer to see quality first remains large (22 percentage points) and significant in regression analyses that control for treatment, gender, age and for whether advisors faced a conflict of interest in the main experiment and whether they were assigned to their preferred order in the main experiment (t -stat= -7.18 ,

$p < 0.001$; see Online Appendix C.2). Altogether, these findings provide support for the interpretation that preferring to see the signal of quality first is a form of moral commitment, which correlates with the take up of a stronger form of commitment: blinding oneself from incentives altogether.

6.2 Explanations for Choice of Information Order

To gather further evidence on whether individuals choose to see quality first to commit to moral judgement, we make use of advisors' self-reported reasons for their choices between information orders collected for a subsample of the Choice Free experiment. The average classification of two independent raters reveals that advisors in the experiment rarely report that they are indifferent between seeing the incentive first or assessing quality first (on average, 10% of the comments), which suggests that indifference is not a main driver of choices. Further, advisors who choose to see the quality signal first are more likely to report doing so to limit bias in their evaluation, as compared to those preferring to see the incentive first (an average of 41% of AMT participants and 53% of professionals versus 5% of AMT participants and 7% of professionals in the two treatments, respectively, χ^2 -stat = 403.6, $p < 0.001$). These findings are consistent with the interpretation that many advisors anticipate the effect of seeing quality first, and preferred to commit to accurate and therefore moral judgement. Conversely, advisors who choose to see the incentive first report to be interested in the commission (an average of 36% of the cases for both AMT and for professionals) or to be motivated by other reasons (more details in Online Appendix C.2).

Result 2(a). 41% of advisors are willing to pay to see the incentive first, while 30% of advisors are willing to pay to see quality first. Their choices correlate with overall morality, with more selfish advisors being more likely to prefer to see the incentive first, and with preferences for blinding.

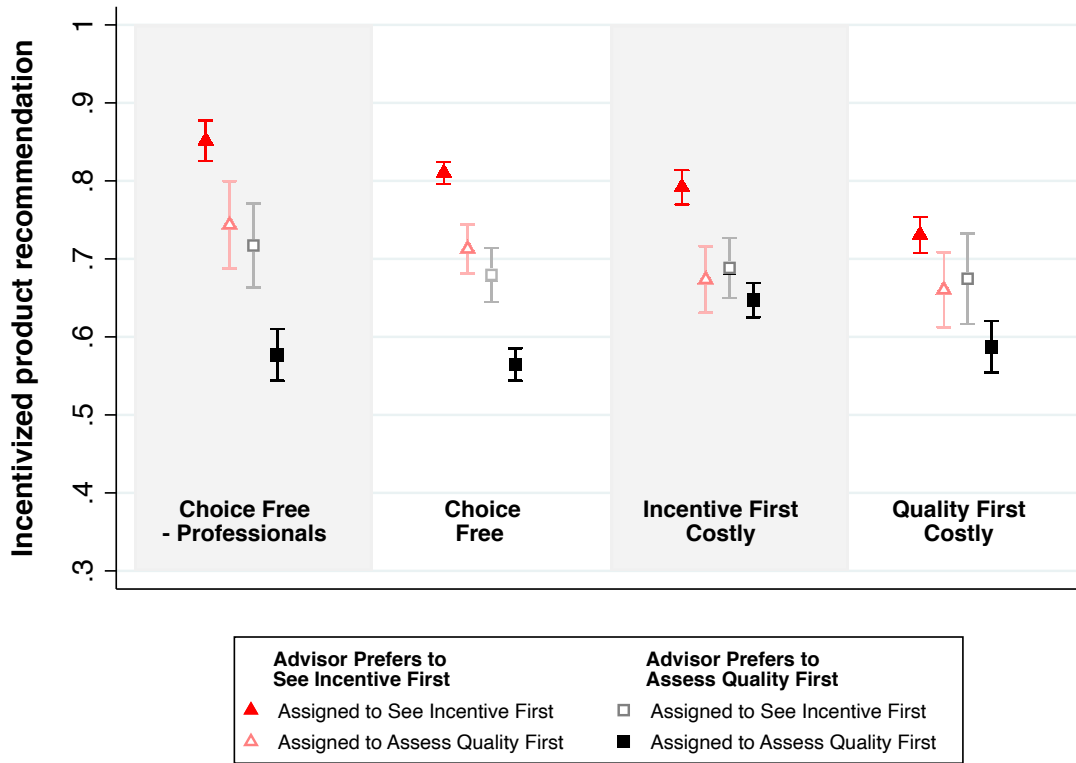
7 Does Demanding Flexibility or Commitment Affect Advice?

Given the heterogeneity in preferences for information order, a central question is how choosing a particular information order affects recommendations. What is the effect of *experiencing* commitment or flexibility?

Figure 6 displays advisors’ recommendation decisions conditional on their preference for and assignment to an information order, focusing on cases in which there is a conflict between the signal of quality about product B and the advisor’s incentive (in Online Appendix C.2 we also provide the figure for cases in which there was no conflict). For advisors who are assigned their preference, recommendation decisions are significantly different depending on the information order. Across all treatments, advisors who prefer and are assigned to see the incentive first (left-most triangle in each cluster in Figure 6) recommend the incentivized product at highest rate. By contrast, those who prefer and are assigned to see quality first (right-most square in each cluster in Figure 6), recommend the incentivized product significantly less often in all cases (t -test, all $p < 0.001$). These results are confirmed by the regression analysis reported in Table 3, where we report coefficient estimates of a linear probability model of the advisor’s decision to recommend the incentivized product for advisors who are assigned their preferred order (column (1)) and those who are not (column (2)), and all together (column (3)). If advisors are assigned their preference, those who prefer to see the incentive first are 19.5-percentage-points more likely to recommend the incentivized product than those who prefer to see quality first (t -stat = 12.17, $p < 0.001$). There is no difference for advisors who do not receive their preferred order. These results reveal that differences in recommendation are not only due to sorting and that *experiencing* information in the desired order is central to the ability to provide self-serving recommendations or constrain them.

In the absence of conflict, advisors are significantly more likely to recommend the incentivized option, and the difference between advisors who prefer to see the incentive first and those who prefer to see quality first is significantly smaller. Overall, advisors exhibit a preference for recommending product A, despite the absence of a conflict of interest. Despite this preference, the difference in recommendations between advisors who prefer to see the incentive first and those who prefer to see quality first remains qualitatively similar focusing on cases in which the incentive is to recommend product A or product B, as shown in separate regressions in Online Appendix C.2.

To examine whether actively choosing an information order that provides more cognitive flexibility could reduce the scope for rationalizing self-serving behavior, we conduct two sets of analyses. First, we investigate whether advisors who prefer to see the incentive first are more likely to recommend the incentivized product when they are assigned to see information in their desired order. On average, advisors who



Notes: This figure presents the covariate-adjusted recommendations of the incentivized product when there is a conflict between the signal of quality and the advisor’s incentive, with the same covariates as in Figure 4. Error bars indicate ± 1 SE.

Figure 6: Advisor Recommendations

choose to see the incentive first are 9.8 percentage points more likely to recommend the incentivized product if they are assigned their preferred order (t -stat= 3.66, $p < 0.001$). This evidence indicates that, even if individuals’ actively choose to have more cognitive flexibility, they still benefit from experiencing it.

Second, we compare the size of the gap in recommendations between advisors who choose flexibility or commitment and are assigned their preference to the gap in recommendations observed in the NoChoice experiment, where individuals are randomized to a given information order. To compare the two experiments, we focus on the Choice Free treatment conducted on AMT, since it has the same incentives and sample of the NoChoice experiment. In the Choice experiment, we estimate a 23.5pp gap, which is not significantly different from the gap estimated in the NoChoice experiment (t – stat = 1.26, $p = 0.207$), but directionally larger by about

Table 3: Advisor Recommendations

	(1)	(2)	(3)
	Recommend incentivized product		
<i>Assignment:</i>	Assigned Pref.	Not Assigned Pref.	Both
Prefer to See Incentive First	0.195*** (0.016)	0.003 (0.029)	0.181*** (0.015)
Not Assigned Preference			0.060*** (0.021)
Prefer to See Incentive First X Not Assigned Preference			-0.140*** (0.026)
No Conflict	0.256*** (0.020)	0.202*** (0.033)	0.236*** (0.018)
No Conflict X Prefer to See Incentive First	-0.137*** (0.025)	0.012 (0.045)	-0.098*** (0.022)
No Conflict X Not Assigned Preference			0.019 (0.025)
Choice Free–Professionals	-0.026 (0.025)	0.051 (0.044)	-0.006 (0.022)
See Incentive First Costly	0.035** (0.017)	0.020 (0.031)	0.031** (0.015)
Assess Quality First Costly	0.004 (0.030)	0.093* (0.052)	0.027 (0.026)
Incentive for B	-0.171*** (0.013)	-0.187*** (0.023)	-0.175*** (0.011)
Female	0.005 (0.013)	-0.015 (0.023)	-0.001 (0.011)
Age	-0.002*** (0.001)	-0.003*** (0.001)	-0.002*** (0.000)
Constant	0.737*** (0.027)	0.864*** (0.048)	0.755*** (0.025)
Observations	4448	1460	5908
R^2	0.106	0.083	0.097

Note: This table displays the estimated coefficients from linear probability models on the advisor’s decision to recommend the incentivized option. Column (1) focuses on individuals who are assigned their preference, while column (2) focuses on individuals who are not assigned their preference. Both groups are merged in column (3). Prefer to See Incentive First is an indicator of the advisor’s preference, and Not Assigned Preference is an indicator for not receiving the preferred order. No Conflict is an indicator for the cases in which the signal of quality is not in conflict with the advisor’s commission. See Incentive First Costly and Assess Quality First Costly are indicator variables that take value 1 in the respective treatment, 0 otherwise. All regression models include individual controls for the advisor’s gender and age, each wave of the experiment, whether incentives were probabilistic, the position of the products on the screen and the interaction between these two variables. The same analysis including a measure of advisor’s selfishness are shown in Online Appendix C. Robust standard errors (HC3) in parentheses. * $p < .10$; ** $p < .05$; *** $p < .01$

7pp.¹⁷

These two sets of analyses show that actively pursuing cognitive flexibility, by choosing to see the incentive first, does not fully remove advisors’ ability to leverage that information order to their advantage to make self-serving recommendations,

¹⁷We thank a reviewer for suggesting this comparison. We should note that the NoChoice experiment has a smaller sample than the Choice experiment and, as a result, has wider confidence intervals (6-28pp), which overlap with the more precise estimate obtained in the Choice experiment (19-28pp). We provide a detailed comparison of recommendation behavior across these two experiments in Online Appendix C.3.

though it may directionally limit it.

We also examine whether pursuing commitment, by choosing to see quality first, is an effective strategy for preventing self-serving behavior. Our results reveal that it is: Conditional on preferring to see quality first, those actually assigned to assess quality first are less likely to make the incentivized recommendation. Relative to advisors who are assigned to receive cognitive flexibility, those who are assigned moral commitment are 9pp less likely to recommend the incentivized product ($t - \text{stat} = 3.05$, $p = 0.002$). This result suggest that limiting self-serving behavior requires temporarily blinding these individuals from receiving information on their incentive.

Result 2(b). Advisors who choose and are assigned to see the incentive first are significantly more like to recommend the incentivized product than advisors who choose and are assigned to see quality first. When advisors are not assigned their preferred information order there is no significant difference in recommendations.

7.1 Evidence of Belief Distortion

To examine whether advisors exhibit biases in belief updating after pursuing and getting flexibility or commitment, we study how individuals update their beliefs from the prior of 0.50 after seeing the signal of quality. For this analysis we merge the beliefs of all advisors in the Choice experiment and follow the approach of Möbius et al. (2022) to examine belief updating relative to Bayes' Rule. For this purpose, we use the continuous belief measure (0-100) that we elicit after our incentivized belief measure (which is in bins). In Online Appendix C.2 we report the analyses that leverage the incentivized belief measure showing qualitatively similar results.

We test whether belief updating about the signal of quality among advisors who prefer and are assigned to see the incentive first differs from that of those who prefer and are assigned to see the quality signal first.¹⁸ In the experiment, the advisor could get a signal that was in conflict with their incentive ($\sigma = c$) or one that was aligned with her incentive ($\sigma = nc$). We denote the advisors' posterior belief that the likelihood of product B is low with $\hat{\mu}$. Möbius et al. (2022) show that the relationship between the advisor's logit belief about quality and the Bayesian benchmark can be estimated using a linear model that includes the loglikelihood

¹⁸Beliefs about quality are one of the potential beliefs that individuals distort; others include beliefs about ethicality, which we did not measure in the experiment.

ratio of each possible signal. We denote as γ_C the log likelihood ratio of a signal in conflict with the incentive and γ_{NC} the log likelihood ratio of a signal not in conflict with the incentive.¹⁹ Conditional on the advisor’s preference and assignment, we estimate the following model of belief updating:

$$\text{logit}(\hat{\mu}) = \beta_C \cdot I\{\sigma = c\} \cdot \gamma_C + \beta_{NC} \cdot I\{\sigma = nc\} \cdot \gamma_{NC} + \epsilon_i$$

where the parameters β_C and β_{NC} indicate the responsiveness of the advisor’s beliefs to a signal in conflict with the incentive or not in conflict with the incentive, respectively, relative to the Bayesian benchmark. If individuals are Bayesian, $\beta_C = \beta_{NC} = 1$.

In Panel A of Table 4, we report estimates of the aforementioned parameters. Column (1) focuses on advisors who are assigned their preference, while column (2) focuses on those who are not assigned their preference. Columns (3) and (4) conduct the same analysis restricting the sample to exclude advisors who update in the wrong direction, from the prior of 0.5, given the signal.

Panel A of Table 4 shows that, similar to Möbius et al. (2022), beliefs exhibit conservatism, as all coefficients are significantly smaller than 1. In our aggregate sample we find evidence for a directional bias in updating: column (1) shows that advisors are more responsive to signals that are not in conflict with the incentive ($\beta_{NC} = 0.380$) than to signals in conflict with the incentive ($\beta_C = 0.307$, F – stat = 5.57, $p = 0.018$). The estimated parameters are similar for the case in which advisors are not assigned to their preferences, though the estimates are less precise and therefore the difference is not statistically significant. Although there is higher responsiveness to signals when advisors who update in the wrong direction are excluded (columns (3) and (4)), the gap between signals in conflict and not in conflict with the incentives persists (F – stat = 12.06, $p < 0.001$). This finding is in line with prior work suggesting that individuals update less when facing negative news (e.g., Eil and Rao, 2011; Möbius et al., 2022).

To study whether individuals who pursue and get to receive information about their incentive first exhibit more distorted beliefs, both in the form of conservatism and directional bias, we estimate the model separately for advisors who prefer order $f \in \{i, q\}$. The results are reported in Panel B of Table 4. We find evidence that

¹⁹In our experiment, when the signal was a blue ball, we have $\gamma = -\log(2)$, when the signal is red, we have $\gamma = \log(3)$. Whether these likelihood ratios are considered conflict or no conflict depends on whether the commission was for product A or B.

Table 4: Belief Updating

	(1)	(2)	(3)	(4)
	Log-odds Belief			
<i>Assignment:</i>	Assigned Pref.	Not Assigned Pref.	Assigned Pref.	Not Assigned Pref.
<i>Data:</i>	All	All	Excl. update in wrong direction	Excl. update in wrong direction
Panel A: Pooled				
β_C	0.305*** (0.016)	0.312*** (0.028)	0.549*** (0.014)	0.575*** (0.024)
β_{NC}	0.380*** (0.027)	0.378*** (0.046)	0.644*** (0.023)	0.646*** (0.038)
Panel B: By Choice of Information Order				
$\beta_C^{f=i}$	0.267*** (0.022)	0.299*** (0.038)	0.525*** (0.019)	0.567*** (0.033)
$\beta_C^{f=q}$	0.346*** (0.023)	0.327*** (0.040)	0.574*** (0.020)	0.583*** (0.035)
$\beta_{NC}^{f=i}$	0.324*** (0.038)	0.405*** (0.067)	0.626*** (0.033)	0.677*** (0.055)
$\beta_{NC}^{f=q}$	0.444*** (0.039)	0.347*** (0.063)	0.664*** (0.033)	0.609*** (0.054)
Observations	4385	1447	3674	1193
$\beta_C^{f=q} = \beta_C^{f=i}$	0.014	0.613	0.078	0.743
$\beta_{NC}^{f=q} = \beta_{NC}^{f=i}$	0.029	0.533	0.417	0.374

Notes: The outcome in all regressions is the log belief ratio. β_C^f and β_{NC}^f are the estimated effects of the log likelihood ratio for conflict and no conflict signals, respectively, for advisors who prefer order $f = i$ indicates a preference to see the incentive first, and $f = q$ indicates a preference to see quality first). Columns (1) and (2) include all advisors. Columns (3) and (4) exclude advisors who updated in the wrong direction. Columns (1) and (3) include only advisors who were assigned their preference, while columns (2) and (4) include only advisors who were not assigned their preference. Robust standard errors (HC3) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

the order of information affects belief distortion. Column (1) of Panel B shows that, for advisors who receive information in their desired order, seeing the incentive first (as opposed to quality first) leads to a lower responsiveness to signals in conflict with the incentive ($\beta_C^{f=i} = 0.267$ versus $\beta_C^{f=q} = 0.346$, $t - \text{stat} = 2.45$, $p = 0.014$). A decrease in responsiveness to signals is also observed when advisors see a signal that is not in conflict with the incentive ($\beta_{NC}^{f=i} = 0.324$ while $\beta_{NC}^{f=q} = 0.444$, $t - \text{stat} = 2.19$, $p = 0.029$). When we exclude advisors who update in the wrong direction, we find that seeing the incentive first leads to a directionally larger and significant decrease in attention to signals in conflict with the incentive ($t - \text{stat} = 1.76$, $p = 0.078$), and a smaller directional decrease in response to signals that are not in conflict with the incentive ($t - \text{stat} = 0.81$, $p = 0.417$). Notably, these differences in updating do not appear when advisors are not assigned to receive information in their desired order, as displayed in columns (2) and (4). These findings are broadly in line with the theoretical framework, as they show that advisors pay less attention to signals when

the incentive is seen first, particularly when these conflict with the incentive.²⁰

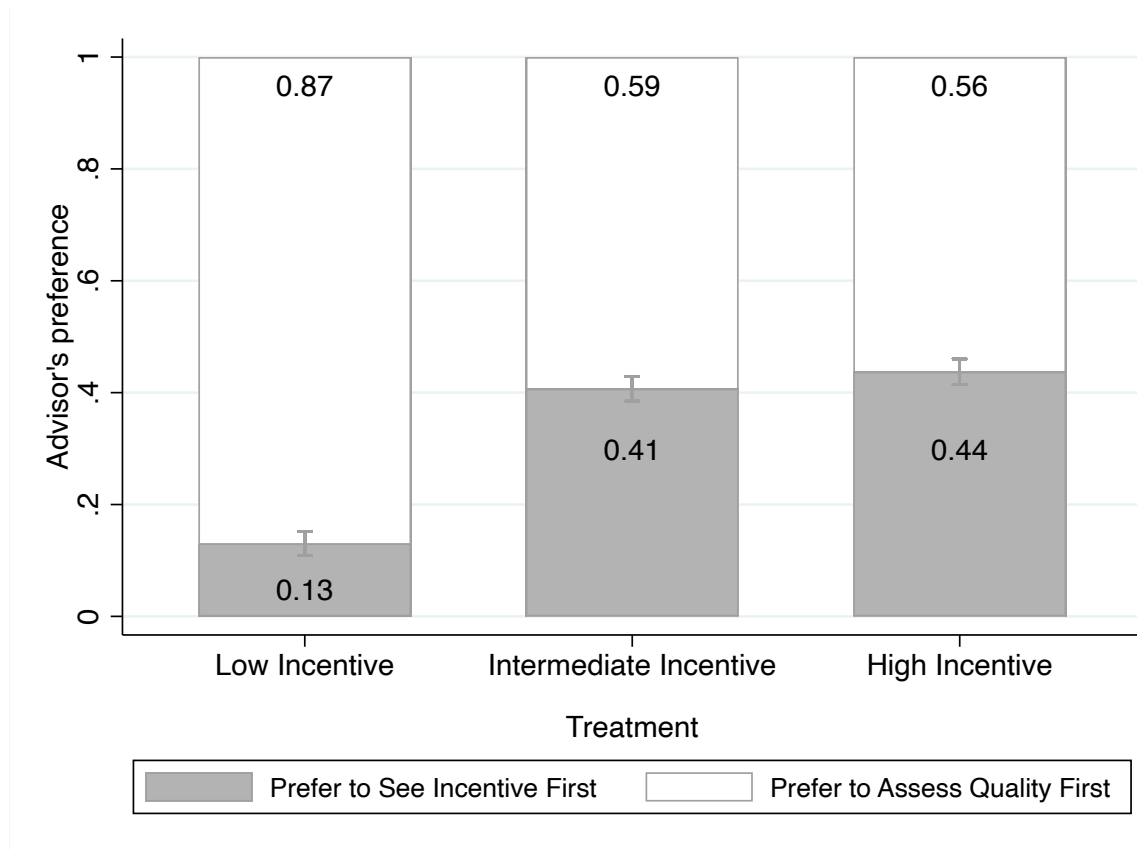
8 Additional Tests of Sophistication

8.1 Advisors' Preferences and Incentives

In the ChoiceStakes experiment, we test whether advisors' demand to see the incentive first responds to the financial gain from recommending the incentivized product. If the gains from recommending the incentivized product decrease, advisors have a smaller incentive to distort their beliefs, making the demand for cognitive flexibility (seeing the incentive first) less desirable. Figure 7 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. 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 = 0.96, $p = 0.337$), despite the fact that the commission is doubled. These results are confirmed in regression analyses in Online Appendix C.6.

We conduct exploratory analyses on advisors' recommendations of the incentivized product in each treatment, shown in Online Appendix C.6. We pool all treatments together, and test whether, when assigned to their preferred information order, advisors who prefer to see the incentive first are more likely to recommend the incentivized product. When advisors are assigned their preferred order, they are 14 percentage points more likely to recommend the incentivized product. However, when not assigned to their preferred order, advisors who expressed a preference to see the incentive first are no more likely to recommend the incentivized product than advisors who indicated the opposite preference.

²⁰In Online Appendix C we separate the analysis by signal, and find that the decrease in responsiveness to signals in conflict is strong for signals of \$0, whereas for signals of \$2 we see a directional decrease in responsiveness both for signals of conflict and for signals of no conflict. The difference in updating patterns between the two signals could arise from the differences between product A and B . Whereas justifying recommendations of product B required advisors to dismiss "bad news" about the quality of product B – a \$0 signal –, justifying recommendations of product A following a \$2 signal did not necessarily require them to dismiss positive signals about the quality of product B . To rationalize recommendations of product A , advisors could have used other justifications, such as the fact that the quality of B was uncertain whereas the quality of A was certain. This potential explanation is in line with our findings of substantially stronger preferences for A in our experiment even when advisors did not face a conflict of interest.



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 7: Advisor’s Preference to See Incentive First, by Treatment

Most models of motivated cognition assume that belief distortion is driven by incentives (e.g., Bénabou and Tirole, 2011; Brunnermeier and Parker, 2005), yet some evidence suggests that sometimes belief distortion is insensitive to stakes (e.g., Coutts, 2019; Engelmann et al., 2019). This experiment shows that advisors’ preferences to see the incentive first respond to incentives to recommend the incentivized product, in line with our theoretical framework and other models of motivated beliefs (e.g., Bénabou and Tirole, 2011; Brunnermeier and Parker, 2005). When doubling the commission of the advisor, however, the preference to see the incentive first increases by only 3 percentage points, less than 10 percent. Our experiment thus suggests that demand for cognitive flexibility increases concavely with the incentive to recommend the incentivized product. This evidence can be useful for further theoretical and empirical work on self-deception to better understand the role of

incentives in belief distortion.

8.2 Do Third Parties Anticipate the Effect of Information Sequence on Advisor’s Behavior?

To better understand the motives driving advisors’ preferences for information order, we investigate whether third parties anticipate the effect of information order. In the Information Architect-Experiment we focus on choices of information order by information architects (IAs) who have incentives that are either aligned with those of the advisors (*IA-Advisor*) or with those of the client (*IA-Client*). Our findings show that, in the *IA-Advisor* treatment, the fraction of IAs who choose for the advisor to see their incentive first is significantly larger than in the *IA-Client* treatment, where advisors’ incentives are aligned with the client (58% vs 44%, $N = 498$, Z – stat = 3.23, $p = 0.001$), and this difference is robust to controlling for demographics (see Online Appendix C.7). These findings are suggestive that third parties anticipate the effect of information order on behavior. We further find that the fraction of IAs who chose for the advisor to see the incentive first in the *IA-Advisor* treatment is similar to the average fraction of advisors who prefer to see the incentive first in the *Choice Free* treatment of the Choice experiment (56%) (Z – stat = 0.497, $p = 0.62$). Since IAs did not receive any information about the realized incentive and quality signal, this result suggests that choices to see the incentive first in the Choice Free treatment are not entirely explained by individuals choosing to see the incentive first to satisfy curiosity.

9 Conclusion

A large body of research has shown that self-serving behavior becomes more likely when individuals can distort their beliefs in order to preserve their self-view as moral. Yet, there are cognitive constraints to the ability to distort beliefs in presence of informative signals. In this paper we ask whether individuals actively take action to constrain their ability to distort beliefs, a form of commitment to moral behavior, or rather seek out the cognitive flexibility needed to distort beliefs, and investigate how, conditional on preferences, being assigned to *experiencing* commitment or flexibility affect self-serving behavior.

We find that a sizable fraction of advisors (30-45%) is willing to take up an oppor-

tunity to constrain belief distortion by seeing quality information first, even when this choice is costly. These preferences are correlated with the take-up of stronger forms of moral commitment and with advisors' morals, measured by their choices when a conflict of interest is always present. We further document that advisors who preferred commitment—wanting to first assess the quality of the product—but that were assigned to first learn about their incentives instead, were more likely to provide biased recommendations than advisors who actually got to see quality information first. This finding suggests that actively *wanting* to commit to unbiased (and moral) judgment may not be enough to prevent self-serving recommendations when the environment in which advisors make decisions is structured in a way that amplifies cognitive flexibility. *Experiencing* commitment is important for reducing the extent of self-serving behavior.

Alongside the preference for moral commitment expressed by some advisors, we find that a considerable share of advisors (40-55%) actively seek out cognitive flexibility by asking to see their incentive before making quality assessments, even when doing so is costly. Actively seeking such cognitive flexibility does not entirely preclude individuals from being able to distort their beliefs. Conditional on preferring to see the incentive first, advisors who experienced cognitive flexibility were more likely to make self-serving recommendations than those who did not. These results contribute to the philosophical discussion on the intentionality of self-deception, by showing that individuals can intend to distort beliefs for self-serving reasons and still be successful at doing so. Altogether, our findings suggest that at least a portion of individuals can anticipate some cognitive constraints to belief distortion, suggesting some level of sophistication about their ability to distort their beliefs when potentially inconvenient information cannot be avoided.

Experts across professions are often called to make partially subjective judgments and could be influenced by a variety of incentives. Such incentives can vary in size, ranging from receiving free gadgets like pens or mugs, or meals, to expensive vacations and large payments or commissions (see, e.g, Campbell et al., 2007; Susman, 2008), but they can also be less tangible (e.g, hiring a candidate for reasons other than their qualifications, using information other than merit, such as the authors' names, to evaluate the quality of a research proposal). In all these examples, incentives may sway experts to make less than objective judgments. In our experiments, we mimic such conflict of interests using small monetary incentives. We find that such small incentives can bias judgement and recommendations, leading some ad-

visors to seek out commitment. When experts make a judgment for the first time, being blind to their incentives could significantly affect their evaluations of quality. Although, experts may later on learn about their incentives, the quality judgment formed initially could affect evaluations later on, as suggested by Chen and Gesche (2017) who document the long-lasting effects of first impressions in the domain of financial advice. Whether the effects documented in this paper apply to settings where experts face substantially higher incentives or less tangible incentives than the ones we used in our experiments is an empirical question that could be investigated in future work.

Altogether, our research suggests that how information provision is structured plays an important role in determining the extent of bias in evaluations, and that a proportion of individuals is willing to temporarily blind themselves from potentially biasing information to ensure the fair and moral behavior. Existing work with hiring managers and academic reviewers provides suggestive evidence in line with our findings. For instance, a vast majority of reviewers support double-blind peer review (Yankauer, 1991; Regehr and Bordage, 2006), but demand for double-blind review is quite limited among authors, especially those who they work at less prestigious institutions (McGillivray and De Rainieri, 2018). In the domain of hiring, although some studies report very high take up of blinding in mock up hiring tasks (e.g., 91.3% in Fath, Larrick, and Soll, 2022), such policies are rare in organizational settings (Bortz, 2018). This evidence could reflect the heterogeneity of preferences we document in our experiment.

The findings in this paper can have important implications for the design of expert systems, suggesting that both organizational design and the selection of experts into organization may occur with commitment or flexibility goals in mind. Organizations often have autonomy and discretion over how to design the information that is presented. The information structure an organization ultimately implements is important, as experiencing commitment or flexibility can alter the extent of self-serving behavior in organizations.

Reference List

- Abeler, J., Nosenzo, D., and Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, 87(4), 1115–1153.
- Anderson, N. H. (1965). Primacy effects in personality impression formation using a generalized order effect paradigm. *Journal of personality and social psychology*, 2(1), 1.
- Anderson, N. H., and Barrios, A. A. (1961). Primacy effects in personality impression formation. *The Journal of Abnormal and Social Psychology*, 63(2), 346–350.
- Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology*, 41, 258–290.
- Babcock, L., Loewenstein, G., Issacharoff, S., and Camerer, C. (1995). Biased judgments of fairness in bargaining. *The American Economic Review*, 85(5), 1337–1343.
- Barfort, S., Harmon, N.A., Hjorth, F., and Olsen, A.L. (2019). Sustaining Honesty in Public Service: The Role of Selection. *American Economic Journal: Economic Policy*, 11(4), 96–123.
- Bénabou, R. (2013). Groupthink: Collective delusions in organizations and markets. *Review of Economic Studies*, 80(2), 429–462.
- 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. (2015). The economics of motivated beliefs. Jean-Jaques Laffont Lecture, *Revue d'Économie Politique*, 125(5), 665–85.
- Bénabou, R., and Tirole, J. (2016). Mindful Economics: The production, consumption and value of beliefs. *Journal of Economic Perspectives*, 30(3), 141–64.
- Bénabou, R., Falk, A., and Tirole, J. (2018). Narratives, Imperatives and Moral Reasoning. NBER Working Paper #24798.
- 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.
- Bortz, D. (2018). Can Blind Hiring Improve Workplace Diversity? *HR Magazine*, March 20, 2018. Society for Human Resource Management.
- Brunnermeier, M.K., and Parker, J.A. (2005). Optimal expectations. *American Economic Review*, 95(4), 1092–1118.
- Camerer, C.F., Hogarth, R.M. (1999). The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework. *Journal of Risk and Uncertainty* 19, 7–42.
- Campbell, E. G., Gruen, R. L., Mountford, J., Miller, L. G., Cleary, P. D., and Blumenthal, D. (2007). A national survey of physician – industry relationships. *New England Journal of Medicine*, 356(17), 1742–1750.
- 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.
- Coutts, A. (2019). Testing models of belief bias: An experiment. *Games and Economic Behavior*, 113, 549–565.
- Chen, Z. and Gesche, T. (2017). Persistent bias in advice-giving. Mimeo.
- 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 Vivaldi, E. (2019). Predicting science to improve science. *Science*, 366(6464), 428–429.
- DeJong, C., Aguilar, T., Tseng, C. W., Lin, G. A., Boscardin, W. J., and Dudley, R. A. (2016). Pharmaceutical industry-sponsored meals and physician prescribing patterns for Medicare beneficiaries. *JAMA internal medicine*, 176(8), 1114–1122.
- 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.
- Engelmann, J., Lebreton, M., Schwardmann, P., van der Weele, J. J., and Chang, L. A. (2019). Anticipatory anxiety and wishful thinking. *Mimeo*.
- Enke, B., Gneezy, U., Hall, B., Martin, D., Nelidov, V., Offerman, T., and van de Ven, J. (2021). Cognitive Biases: Mistakes or Missing Stakes? *The Review of Economics and Statistics*, 1–45.
- 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*, 87 (5), 2205–29.

- Fath, S., Larrick, R.P., and Soll, J.B. (2022). Blinding curiosity: Exploring preferences for “blinding” one’s own judgment. *Organizational Behavior and Human Decision Processes* 170, 104135.
- 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.
- Gill, A., Heinz, M., Schumacher, H., and Sutter, M. (2022). Trustworthiness in the Financial Industry. CEPR Discussion Paper No. DP15147.
- 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. (2005). Deception: The Role of Consequences. *American Economic Review* 95 (1), 384-94.
- 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. (2017). 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.
- Hanna, R., and Wang, S. (2017). Dishonesty and Selection into Public Service: Evidence from India. *American Economic Journal: Economic Policy* 9 (3), 262–90.
- Hsee, C.K. (1996). Elastic justification: How unjustifiable factors influence judgments. *Organizational Behavior and Human Decision Processes*, 66(1), 122–129.
- Huber, C., Huber, J. (2020). Bad bankers no more? Truth-telling and (dis) honesty in the finance industry. *Journal of Economic Behavior Organization*, 180, 472–493.
- 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. *Judgment and Decision Making*, 8, 407–424.
- Kahneman, D. (1973). *Attention and Effort*. Prentice-Hall, USA.
- Karlsson, N., Loewenstein, G., and Seppi, D. (2009). The ostrich effect: Selective attention to information. *Journal of Risk and Uncertainty* 38, 95–115.
- 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.
- Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (1997). Motivated attention: Affect, activation, and action. *Attention and orienting: Sensory and motivational processes*, 97, 135.
- 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.
- Marechal, M.A., and Thoni, C. (2019). Hidden Persuaders: Do Small Gifts Lubricate Business Negotiations? *Management Science* 65 (8), 3470–3469.
- McGillivray, B., and De Ranieri, E. (2018). Uptake and outcome of manuscripts in Nature journals by review model and author characteristics. *Research Integrity and Peer Review* 3 (5).
- 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.
- Möbius, M., Niederle, M., Niehaus, P. and Rosenblat, T. (2022). Managing self-confidence: Theory and experimental evidence. *Management Science*, forthcoming.
- 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.
- Pace, D., and van der Weele, J. (2021). Fair shares and selective attention. Tinbergen Institute Discussion Paper 2021-066.
- 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.
- Regehr, G., and Bordage, G. (2006). To blind or not to blind? What authors and reviewers prefer. *Medical education* 40 (9), 832–839.
- Robertson, C.T., and Kesselheim, A.S. (2016). *Blinding as a Solution to Bias: Strengthening Biomedical Science, Forensic Science and Law*. Elsevier, USA.
- 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. (2021). Self-Persuasion: Evidence from Field Experiments at Two International Debating Competitions. *American Economic Review*, forthcoming.
- Schwartzstein, J. (2014). Selective attention and learning. *Journal of the European Economic Association* 12(6), 1423–1452.
- Serra, D., Serneels, P., and Barr, A. (2011). Intrinsic motivations and the non-profit health sector: Evidence from Ethiopia. *Personality and Individual Differences*, 51(3), 309–314.
- Serra-Garcia, M., and Szech, N., (2021). The (in) elasticity of moral ignorance. *Management Science*, forthcoming.
- 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 Haggmayer, 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.

- Susman, T. (2008). Private Ethics, Public Conduct: An Essay on Ethical Lobbying, Campaign Contributions, Reciprocity and the Public Good. *Stanford Law Policy Review*, 19(1):10-22.
- Tasoff, J., and Madarasz, K. (2009). A model of attention and anticipation. Working paper.
- Tetlock., P.E. (1983). Accountability and the perserverence of first impressions. *Social Psychology Quarterly*, 285-292.
- Trivers, R. (2011). *The Folly of Fools: The Logic of Deceit and Self-Deception in Human Life*. Basic Books.
- Tversky, A., Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, 185(4157), 1124-1131.
- Yankauer, A. (1991). How blind is blind review? *American Journal of Public Health* 81 (7), 843–845.
- Yates, J. F., Curley, S. P. (1986). Contingency judgment: Primacy effects and attention decrement. *Acta psychologica* , 62(3), 293-302.
- Zimmerman, F. (2018). The dynamics of motivated beliefs. *American Economic Review*, 110(2), 337–61.

Online Appendix for
Cognitive Flexibility or Moral Commitment?
Evidence of Anticipated Belief Distortion

Silvia Saccardo and Marta Serra-Garcia

May 2022

Table of Contents

A	Theoretical Model: Proofs, Additional Results and Example	4
A.1	Main Model: Effects of Information Order	4
A.2	Main Model: Preferences for information order	5
A.3	Incentives and preferences for information order	8
B	Detailed Experimental Design and Procedures	10
B.1	The Experiments	10
	Table B.1 - The Experiments	10
B.2	Sample, Recruitment Procedures and Exclusion Criteria	12
B.3	Additional measures	14
B.4	Exclusion Criteria	16
C	Additional Analyses	17
C.1	NoChoice Experiment	17
	Table C.1 - Recommendations	17
	Table C.2 - Recommendations including Inattentive	18
C.2	Choice experiment	19
	C.2.1 Preferences	19
	Table C.3 - Preferences for Information Order by Wave	19
	Table C.4 - Preferences for Blindness and Preferences for Information Order	19
	Table C.5 - Preferences for Blindness, Information Order & Selfishness . . .	20
	C.2.2 Recommendations	21
	Figure C.1 - Advisor Recommendation - No Conflict	21
	Table C.6 - Advisor Recommendations	22
	Table C.7 - Advisor Recommendations: Incentive for A	23
	Table C.8 - Advisor Recommendations: Incentive for B	24
	C.2.3 Beliefs	25
	Figure C.2 - Beliefs - Get Preferred Information Order	25
	Figure C.3 - Beliefs - Do Not Get Preferred Information Order	25
	Table C.9 - Belief Updating when Signal is \$0	26
	Table C.10 - Belief Updating when Signal is \$2	27
	Table C.11 - Belief Updating: Correct Choice	28
	C.2.4 Explanations for Choices	29
	Table C.12 - Advisors' Explanations: Detailed Results	29
C.3	Comparing the Choice and NoChoice Experiments	29
	Table C.13 - Advisor Recommendations	31

Table C.14 - Recommendations in the NoChoice and Choice Experiment . . .	32
C.4 The Higher Incentives Treatments	33
Table C.15 - Preference for Information Order: Including Incentives Treatments	34
Table C.16 - Advisor Recommendations: Including Incentives Treatments . .	35
C.5 Including Inattentive Participants	36
Table C.17 - Preference for Information Order—Including Inattentive	36
Table C.18 - Advisor Recommendations—Including Inattentive	37
C.6 The ChoiceStakes Experiment: Additional Results	38
Table C.19 - Preference for Information Order	38
Table C.20 - Advisor Recommendations	39
C.7 The Information Architect Experiment: Additional Results	40
Table C.21 - IA Preferences by condition	40
D The NoChoiceSimultaneous Experiment	41
D.1 Experimental Design	41
D.2 Results	42
Table D.1 - Advisor Recommendations - No Choice (Simultaneous)	43
E The Choice Deterministic Experiment	44
E.1 Experimental Design	44
E.2 Results	45
Figure E.1 - Recommendations	46
Table E.1 - Recommendations: Assigned Preferences	47
F Additional Data: Predictions	48
Figure F.1 - Predicted and Actual Effect of Seeing the Incentive First on Rec- ommendations	49
G Experimental Instructions	50
G.1 Choice Experiment	50
G.2 Information Architect experiment	65

A Theoretical Model: Proofs, Additional Results and Example

A.1 Main Model: Effects of Information Order

We discuss the strategies and equilibria when the advisor is exogenously assigned to see quality first ($f = q$) and when the advisor is exogenously assigned to see the incentive first ($f = i$). Let us denote the game as G^f .

Proposition A.1. *If $m_0 > \iota$, the PBE of G^f , with $f \in \{i, q\}$, is characterized by Self 0 not suppressing the signal of quality. If $m_0 \leq \iota$, in any PBE of G^f ,*

$$p_s^* = \min\left\{\frac{(1 - \lambda^f)(\iota - \phi M)}{\lambda^f \phi (M - \iota)}, 1\right\}$$

Hence, suppression is more likely when the advisor sees the incentive first.

Proof. The expected utility of Self 0 given p_s is:

$$E(U_0) = \lambda^f \left((1 - \phi)\iota + \phi((1 - p_s)((\iota - m_0)\frac{\iota}{M}) + p_s(\iota - m_0)q) \right) + (1 - \lambda^f)(\iota - \phi m_0)q.$$

Self 1 recommends the incentivized option if she receives $\hat{\sigma} = \emptyset$ with certainty if $\frac{\iota}{Mr(\emptyset)} \geq 1$. Self 1 is uncertain about whether the signal of quality is empty because of Self 0 suppressed it or because it was not encoded to begin with. Using Bayes' Rule,

$$\frac{(p_s \lambda^f + (1 - \lambda^f))\phi}{\lambda^f p_s \phi + (1 - \lambda^f)} \leq \frac{\iota}{M}$$

which implies

$$p_s \leq \frac{(1 - \lambda^f)(\iota - \phi M)}{\lambda^f \phi (M - \iota)}$$

Hence, a selfish Self 0 prefers to suppress as often as possible and hence chooses,

$$p_s^* = \min\left\{\frac{(1 - \lambda^f)(\iota - \phi M)}{\lambda^f \phi (M - \iota)}, 1\right\}$$

Since p_s^* is decreasing with λ^f ,

$$\frac{\partial p_s^*}{\partial \lambda^f} = \frac{\iota - \phi M}{\phi(M - \iota)} \frac{-\lambda^f - (1 - \lambda^f)}{(\lambda^f)^2} < 0,$$

it follows that suppression is more likely when the incentive information is shown first,

than when the signal of quality is shown first.

Ex-ante, if Self 0 is selfish, her expected payoff in G^f is:

$$U_0(G^f) = \lambda^f \left((1 - \phi)\iota + \phi \left((1 - p_s^*) \left((\iota - m_0) \frac{\iota}{M} \right) + p_s^* (\iota - m_0) \right) \right) + (1 - \lambda^f)(\iota - \phi m_0).$$

If Self 0 is moral, her expected payoff in G^f is:

$$U_0(G^f) = \lambda^f \left((1 - \phi)\iota + \phi \left((\iota - m_0) \frac{\iota}{M} \right) \right) + (1 - \lambda^f)(\iota - \phi m_0).$$

We also consider the case in which the advisor is naive, in the sense that she believes that her attention is imperfect under both information orders.

Proposition A.2. *If the advisor believes she does encode quality signals with the same probability for both information orders, she exhibits the same suppression under both information orders. Since encoding is actually less likely when the information is shown first, she still recommends the incentivized product more often when the incentive is shown first.*

Proof. The advisor's belief is $\hat{\lambda}^q = \hat{\lambda}^i = \lambda^i < 1$. Since the advisor's belief is correct when the incentive is shown first, the same prediction holds as for Proposition A.1. When the signal of quality is shown first, Self 0 and Self 1 both believe that there is a probability $1 - \hat{\lambda}^q$ that the signal is not encoded to begin with. Hence, Self 0 suppresses signals that are in conflict with the incentive with probability p_s^* believing the likelihood of encoding a signal is λ^i . Self 1 updates using the same belief about attention. Therefore, the same behavior as in Proposition A.1. would arise. Since in actuality the signal would be encoded less often when the incentive is shown first, there would still be more suppression in that case.

A.2 Main Model: Preferences for information order

Below we provide the proof for Proposition 2 (from the main text).

Proposition 2.

- *If Self 0 is selfish ($m_0 \leq \iota$), she chooses to see the incentive first ($f^* = i$). This order increases the likelihood that Self 1 recommends the incentivized product when the signal is in conflict with the incentive.*
- *If Self 0 is moral ($m_0 > \iota$), she chooses to see quality first ($f^* = q$), which decreases*

the likelihood that Self 1 recommends the incentivized product when the signal is in conflict with the incentive.

Proof. If $m_0 > \iota$, Self 0's utility increases as the likelihood that Self 1 recommends the incentivized product when it is in conflict with the incentive decreases. By choosing $f = q$, the likelihood that the incentivized product is recommended is lowered to $\frac{\iota}{M}$. If $m_0 \leq \iota$, Self 0's utility increases as the likelihood that Self 1 recommends the incentivized product when it is in conflict with the incentive increases. For any p_s^* , the likelihood is higher when $f = i$ because $\lambda^i < \lambda^q$.

The proof of Proposition 3 follows directly from the fact that the advisor does not believe that information order differentially affects her likelihood of encoding a quality signal. Given that she does not anticipate a difference in behavior, she is not willing to pay for any information order.

Numerical Example for NoChoice and Choice. In what follows we use a simple numerical example to illustrate the differences between the predicted behavior of Self 0 and Self 1, when the incentive information is shown first ($f = i$) and when the quality signal is shown first ($f = q$). Consider the case where advisors who see the incentive first encode the quality signal 50% of the time ($\lambda^i = 0.5$), while advisors who see the quality signal first encoded it 70% of the time ($\lambda^q = 0.7$). The incentive ι is 0.15, while the prior likelihood that the signal is in conflict with the incentive is $\phi = 0.4$. The range of moral costs is from 0 to $M = 0.3$.

Given these parameter values, and applying the formula for p_s^* in Proposition A.1., if the incentive is shown first, the optimal likelihood of suppression is

$$p_s^* = \frac{(1 - 0.5)(0.15 - 0.4 \cdot 0.3)}{0.5 \cdot 0.4 \cdot (0.3 - 0.15)} = 0.5$$

Similarly, if the quality signal is shown first, the optimal likelihood of suppression is 0.21. Both if the incentive is shown first and if the signal of quality is shown first, the posterior belief of Self 1 after receiving an empty signal ($\hat{\sigma} = \emptyset$) is 0.5. What differs between both information orders is how often an empty signal is received.

Given Self 0's suppression strategy, what signal distribution does Self 1 receive? We consider a Self 0 who is selfish and suppresses with the likelihoods shown above. We calculate (a) how often Self 1 receives a signal that is in conflict with the incentive, given that the signal was of conflict, and (b) how often Self 1 receives a signal that is not in conflict with the incentive, given that there is no conflict.

Consider the case where the signal of quality is shown first. Then a signal is encoded

with a likelihood of 0.7, and conditional on it being of conflict, it is suppressed with a likelihood of 0.21. Therefore, if there is a conflict with the incentive, Self 1 receives the signal with a 0.55 likelihood ($0.7 \cdot (1 - 0.21)$). If there is no conflict, Self 1 receives the signal with a 0.7 likelihood.

Consider the case where the incentive is shown first. Then a signal is encoded with a likelihood of 0.5, and conditional on it being of conflict, it is suppressed with a likelihood of 0.5. If there is a conflict with the incentive, Self 1 receives the signal with a 0.25 likelihood ($0.5 \cdot (1 - 0.5)$). If there is no conflict, Self 1 receives the signal with a 0.5 likelihood.

Therefore, seeing the incentive first has a (large) first-order effect: it decreases the likelihood that Self 1 learns the actual signal, both if it is in conflict or not in conflict with the incentive. There is a second effect: seeing the incentive first increases the difference in the likelihood that a signal of conflict is received relative to a signal that is not in conflict.

Recommendations would reflect these differences. Consider the case where Self 0s feature low moral costs ($m_0 < i$) and are classified as selfish in 50% of the cases, while Self 0s would exhibit high moral costs in the remaining 50% of the cases. Selfish Self 0s would prefer to see the incentive first, while moral Self 0s would prefer to see quality first. Given their suppression strategies, how often would Self 1 recommend the incentivized product when the signal is in conflict with the incentive? Given the independent draw of moral costs for Self 0 and Self 1, each Self 0 would have a 50% chance to be matched with a selfish Self 1. Consider a selfish Self 0 who chooses to see the incentive first. Then, Self 1 sees a signal in conflict with the incentive with a 0.25 chance, and receives an empty signal with a 0.75. Since Self 1 is selfish with a 0.5 chance, Self 1 would recommend the incentivized product with a 0.875 chance ($0.5 + 0.5 \cdot 0.5 + 0.5 \cdot 0.5 \cdot 0.25$). If, by contrast, Self 0 would choose to see the signal of quality first, and suppress optimally, Self 1 would recommend the incentivized product with a 0.7235 chance ($0.3 + 0.7 \cdot 0.21 + 0.7 \cdot 0.79 \cdot 0.5$).

Consider by contrast a moral Self 0 who chooses to see quality first. Since she prefers not to suppress, Self 1 would recommend the incentivized product with a 0.65 chance ($0.3 + 0.7 \cdot 0.5$). If, by contrast, Self 0 would choose to see the incentive first, due to the lower attention, Self 1 would recommend the incentivized product with a 0.75 chance ($0.5 + 0.5 \cdot 0.5$).

If advisors are sophisticated and can choose their preferred order (Choice Experiment), this numerical example of the model would predict recommendations of the incentivized product to occur in 87.5% of the cases, among those who choose to see the incentive first, and 65% of the cases, among those who choose to see quality first. This 22.5% percentage point gap is similar, though slightly larger, than the gap we observe in the

Choice experiment (19.5% percentage points). Comparing cases in which the advisors receive their preferred information order and cases in which they do not, conditional on preference, this numerical example would predict a 10 to 15 percentage point gap in recommendations, which is close to the 10 to 12 percentage point gap observed in the Choice experiment.

If advisors cannot choose their preferred order, those assigned to see the incentive first would be selfish Self 0s in 50% of the cases, and moral Self 0s in the remaining cases. Combining the behavior of these two types, the numerical example would predict that 81.25% of advisors recommend the incentivized product when the signal is in conflict with the incentive. Among those assigned to see quality first, the numerical example would predict that 68.7% of advisors would recommend the incentivized product when the signal is in conflict with the incentive. This predicted gap of 12.55 percentage points is qualitatively similar, but somewhat smaller than the gap of 17 percentage points observed in NoChoice.

A.3 Incentives and preferences for information order

In the ChoiceStakes experiment, we vary the advisor's incentive ι , while keeping it costly to see the incentive first. Instead of having $\iota = 0.15$, the Low Commission treatment has $\iota = 0.01$ and the High Commission treatment is $\iota = 0.30$. We examine how increasing or decreasing the incentive affects the utility of seeing the incentive first, assuming advisors are sophisticated.

Corollary A.1. *If the advisor's incentive to recommend the incentivized option decreases, the demand to see the incentive first decreases. If the advisor's incentive to recommend the incentivized option increases, the demand to see the incentive first may increase or decrease.*

Proof. When ι decreases, as in the Low Commission treatment, it is more likely that $m_0 > \iota$, and the advisor is more likely to prefer to see quality first ($f = q$). Further, for advisors who still prefer to see the incentive first, the likelihood of suppression decreases when the incentive decreases. Specifically, since $\iota > \phi M$, then p_s^* decreases with ι since:

$$\frac{\partial p_s^*}{\partial \iota} = \frac{1 - \lambda^f (M - \iota) + (\iota - \phi M)}{\lambda^f \phi (M - \iota)^2} > 0.$$

By contrast, if the incentive increases, we have that p_s^* increases. This can increase the utility from seeing the incentive first, as long as $\iota < M$. However, as the incentive becomes higher, it can become higher than the highest moral cost $\iota > M$. Then, the

potential for conflict between Self 0 and Self 1 disappears, and Self 0 no longer strictly prefers to see the incentive first.

B Detailed Experimental Design and Procedures

B.1 The Experiments

Table B.1 reports all the data collected for this paper, their corresponding pre-registration, recruitment platform and incentives for advisors and clients. We pre-registered the design, sample sizes, exclusion criteria, and analyses of all Amazon Mechanical Turk (AMT) experiments on aspredicted.org. The experiment on professionals was not pre-registered. The design of NoChoice experiment and the ChoiceStakes Experiment are described in full in the main text.

Table B.1: The Experiments

Sample-Wave	Aspredicted pre-reg #	Treatment	Advisor/DM Commission	Client's Payoff Balls	Matching with Client	Year	N
Main Text: NoChoice Experiment							
AMT	22709	See Incentive First	\$0.15	\$0 or \$2	1 advisor out of 10	2019	152
		See Quality First	\$0.15	\$0 or \$2	1 advisor out of 10	2019	147
Main Text: Choice Experiment							
AMT-1	23272	Choice Free	\$0.15	\$0 or \$2	1 advisor out of 10	2019	1308
		Incentive First Costly	\$0.15	\$0 or \$2	1 advisor out of 10	2019	1347
AMT-2	42246	Choice Free	\$0.15	\$0 or \$2	1 advisor out of 10	2020	511
		Choice Free	\$15 to 1/100	\$0 or \$20	1 to 1*	2020	542
Professionals	NA	ChoiceFree Professionals	\$15 to 1/100	\$0 or \$20	1 to 1*	2020	712
AMT-3	70817	Choice Free	\$0.15	\$0 or \$2	1 advisor out of 10	2021	213
		Quality First Costly	\$0.15	\$0 or \$2	1 advisor out of 10	2021	1067
		Incentive First Costly	\$0.15	\$0 or \$2	1 advisor out of 10	2021	215
		ChoiceFree Highx10	\$1.5	\$0 or \$20	1 advisor out of 10	2021	275
		ChoiceFree Highx100	\$15	\$0 or \$20	1 to 1	2021	110
Main Text: ChoiceStakes Experiment							
AMT	76771	Low Incentive	\$0.01	\$0 or \$2	1 advisor out of 10	2021	483
		Intermediate Incentive	\$0.15	\$0 or \$2	1 advisor out of 10	2021	511
		High Incentive	\$0.30	\$0 or \$2	1 advisor out of 10	2021	478
Main Text: Information Architect Experiment							
AMT	76771	IA-Advisor	\$0.15	\$0 or \$2	1 advisor out of 10	2021	245
		IA-Client	\$0.15	\$0 or \$2	1 advisor out of 10	2021	253
Appendix: Choice Deterministic							
AMT	82298	ChoiceFree - Replication	\$0.15	\$0 or \$2	1 advisor out of 10	2021	385
AMT		ChoiceFree - Deterministic	\$0.15	\$0 or \$2	1 advisor out of 10	2021	369
Appendix (Additional Exp): NoChoice Simultaneous							
AMT	79521**	See Incentive First	\$0.15	\$0 or \$2	1 advisor out of 10	2021	83
AMT		See Quality First	\$0.15	\$0 or \$2	1 advisor out of 10	2021	70
AMT		Simultaneous	\$0.15	\$0 or \$2	1 advisor out of 10	2021	130
Appendix (Additional Exp): Predictions							
AMT	37081	Prediction	-	\$0 or \$2		2020	288

Notes. This table presents all the experiments we conducted for this paper, their corresponding sample, wave of data collection, pre-registration, treatments, incentive features for advisors and clients in the experiment, matching between advisors and clients, and sample sizes after excluding inattentive participants and participants with inconsistent responses in the MPL measure of moral costs, as pre-registered.

* In these studies, only 1 out of 100 advisors were selected for payment. These advisors were all matched with a client.

** In this study, due to higher rates of inattention than in other studies, we updated the pre-registration to increase the size of the recruited sample, see Aspredicted #82164.

NoChoice Experiment. The NoChoice experiment was conducted on AMT. All details about the experiment are reported in the main text.

Choice Experiment. As displayed in Table B.1, the Choice experiment on Amazon Mechanical Turk (AMT) was conducted in three different waves (AMT-1, AMT-2, AMT-

3). The first wave of the experiment, AMT-1, was conducted in 2019 and randomized participants in the *ChoiceFree* and *See Incentive First treatment*. The second wave of the experiment, AMT-2, was conducted in 2020, and collected additional data for the *ChoiceFree* treatment. As part of this wave of data collection, we randomized whether incentives were identical to those in the first wave of the experiment or probabilistic as in the professional sample. In the treatments with probabilistic incentives, the products were urns containing payoff balls that were worth either \$0 or \$20 (rather than \$0 and \$2 as in the regular treatments), and the commission for the advisor was \$15 to 1 out of 100 participants (instead of paying \$0.15 as in the regular treatments). One out of 100 advisors was selected for payment, and their recommendations were sent to a client. The experiment also counterbalanced whether the incentivized product was presented on the left side or the right side of the screen. These two factors varied independently between subjects (2x2 design). In 2020 we also collected data for the sample of professionals in the *ChoiceFree* treatment (*ChoiceFree-Professionals*). The third wave of the experiment, AMT-3, was conducted in 2021. The majority of the participants recruited for this wave (80%) were randomized into the *ChoiceFree* treatment, the *See Incentive First* treatment and the *See Quality First*. Since we already had data on the latter treatments from prior waves, participants were randomized to those treatments at a 1:1:5 ratio. The remaining 20% of participants was randomized (at a 2:1 ratio) into one of two robustness treatments that increased the incentives for both the advisor and the client incentives in the *ChoiceFree* treatment. The goal of these robustness treatments was to test whether the effects of information order on recommendations documented in the *ChoiceFree* treatment are specific to the small stakes used in the experiment, or whether they persist when advisors conflict of interests that have higher stakes. In the *High Stakes - 10 fold* treatment, we increased the incentives by a factor of 10. We paid each advisor a \$1.50 commission if she recommended the incentivized product. In this treatment, one out of 10 advisors was then matched to a client, who received either \$0 or \$20. In the *High Stakes - 100 fold* treatment, we increased the incentives by a factor of 100, increasing the commission of the advisor to \$15, and matched each advisor with a client, who received either \$0 or \$20.

ChoiceStakes Experiment. In the *ChoiceStakes* experiment, all instructions were identical to those of the *ChoiceFree* experiment, but we varied the size of the commission while keeping the incentives for the clients the same. All details about this experiment are reported in the main text.

Information Architect-Experiment. In the *IA-Experiment*, all instructions were identical to those of the *ChoiceFree* experiment, except that participants were assigned

to the role of Information Architects (IAs). That is, participants were informed that they would be matched with an advisor and a client and that they will have to make a decision about how the advisor receive information. The Instructions for this experiment are reported in Online Appendix G.2. IAs received information about their incentive and were asked to choose an information order for the advisor they were matched with. Importantly, the IA did not receive information about the product that yielded the advisors a commission nor the signal of quality directly, but only determined the order with which advisors receive such information. We subsequently recruited 498 advisors, and presented them with the information order chosen by the IA. We informed these advisors that the order of information was chosen by the IA. Advisors were not informed about IAs' incentives.

B.2 Sample, Recruitment Procedures and Exclusion Criteria

Sample. All experiments were conducted on AMT, except for the ChoiceFree Professionals treatment, which was conducted on Prolific Academic (Palan and Schitter, 2018) and CloudResearch (Litman et al., 2016), to target participants who self-report to work in two industries in which advice is very frequent: finance and insurance, and legal services. 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.

Out of 712 professionals, 677 (95.1%) provided job descriptions that could be used by our independent raters to judge whether their position was fiduciary or not. As mentioned in the text, two independent raters were asked to classify each job title as fiduciary or not fiduciary, based on the description provided by the participant. They were provided the following information regarding what is defined as fiduciary: *“According to Investopedia, a fiduciary is “a person or organization that acts on behalf of another person or persons, putting their clients’ interest ahead of their own, with a duty to preserve good faith and trust. Being a fiduciary thus requires being bound both legally and ethically to act in the other’s best interests. A fiduciary may be responsible for the general well-being of another (e.g. a child’s legal guardian), but often the task involves finances; managing the assets of another person, or a group of people, for example. Money managers, financial advisors, bankers, insurance agents, accountants, executors, board members, and corporate officers all have fiduciary responsibility”.”*

The raters agreed on their classification of fiduciary duty in 89% of the cases (interrater agreement $\kappa=0.85$). In 61.9% of the cases the job title was considered as fiduciary by at least one rater. Focusing on the cases with agreement, 58.0% of the job titles were considered as fiduciary. Job titles frequently found in the data included the word analyst

(financial, actuarial, etc., in 9.4% of the cases), accountant or account manager (11.6% of the cases), and lawyer or paralegal (in 7.2% of the cases). In their job titles, 14% of participants included the word “manager.” Our raters were also asked to classify the job titles into industry (finance and insurance or legal, or neither if it was not clear from the job title). Prolific provides this information for some of our participants, but it was missing in 159 of 496 cases. The agreement between raters regarding industry classification was high for CloudResearch ($\kappa = 0.79$) and somewhat lower for the missing cases on Prolific ($\kappa = 0.66$). Overall, for cases in which there is an agreement (636 out of 712), we find that 72.5% of professionals work in the finance and insurance industry, 18.9% in legal service, and for the remaining 8.7% the industry is unknown.

Recruitment and Procedures. We recruited participants in the role of advisors to a 5-7 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. Professionals and participants in the third wave of the Choice experiment (AMT-3), the IA experiment and the NoChoice-Simultaneous experiment were instead paid \$1 for completing the study. Participants had to be located in the US and have an approval rating of at least 90%.

Upon being recruited, participants were assigned to the role of advisors and, in almost all treatments, informed that one of ten advisors would be matched with a client, as described in Table B.1. Participants were presented with several understanding questions about the products while reading the instructions. Before randomizing participants to treatments, we included one question that participants had to answer correctly in order to continue in the study (i.e., the attention check). Those who failed to answer it correctly, were disqualified from participation and were not randomized to treatments. Advisors were then provided additional information about the experiment, and then moved to a screen where they were given the choice of information order. Advisors were informed that they would receive a commission for recommending one of two products, A or B. In the AMT-3 wave of the Choice experiment, we clarified that the commission was determined at random by the computer. In the Choice Experiments, advisors were prompted to make a choice between information order. A summary of the treatments we ran in each wave is presented in Table B.1. After receiving information about the incentive and the signal of quality, advisors were asked to provide their recommendation to the client. We then collected measures of beliefs, selfishness, and, in wave AMT-3, preferences for blinding.

B.3 Additional measures

In all experiments (except for the Prediction experiment) we collected additional measures.

Beliefs. As explained in the main text, we elicit 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: in most treatments, advisors received \$0.15 for a guess in the correct range. In the ChoiceFree-Professionals and the ChoiceFree-Probabilistic treatment in the AMT-2 wave of the Choice experiment, this payment was \$15 to 1 out of 100 advisors. In the High Stakes - 100 fold treatment, this payment was \$15.

Moral costs. After the belief measure, we asked participants to complete one additional advice task, aimed at measuring advisors’ selfishness/morality using a multiple price list. We informed advisors that they would be asked to make a second recommendation to an advisee –a participant different than the one who received their first recommendation. Advisors were told they would need to make a series of recommendations to another participant (an advisee), choosing between two products, X and Y. Product Y had the same payoffs of product B in the main experiment. 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. Advisors were incentivized to recommend Y, with a \$0.15 commission, and received a signal of quality of Product Y that indicated that a \$0 had been drawn from Y. Given the payoffs of X, recommending Y (the incentivized product) harmed the client if X paid \$2 with a probability of 0.6 or higher. We use this elicitation to measure the advisor’s selfishness, as the number of times the advisor chose to recommend Y, and standardize it within each experiment.

In the ChoiceFree-Probabilistic treatment in the AMT-2 wave of the Choice experiment, these products were scaled up to paying either \$0 or \$20; the commission to the advisor was \$15 to 1 out of 100 participants. At the end of the experiment, we randomly selected one out of 10 advisors, we randomly picked one of the 5 recommendations, and showed them to an advisee. For this purpose, we recruited a total of 866 advisees.

Blinding. In the AMT-3 wave of the choice experiment, we measure preferences for blinding in an additional advice task. In the task, participants learn that they are matched with another participants, a different client from the one of the main task and of the Selfishness task. We present the advisors with two products, 1 and 2, which yield the same expected payoff of products A and B in the main experiment. As in the main exper-

iment, advisor know that, before making their recommendation they will receive a signal about the quality of Product 2. Advisors learn that they will receive a \$0.15 commission depending on their recommendation. The commission can be either for Product 1 or Product 2, determined at random, and advisors are notified that they will learn for which product the commission is before the end of the study. We then ask advisors to choose whether to learn for which product is the commission *before* receiving a signal about the quality of product and making the recommendation, or *after* learning the quality of Product 2 and making the recommendation. That is, in this task, advisors can either learn their incentive before making the recommendation or after the recommendation is made. By choosing the latter option, advisors can ensure that their recommendation is blind to incentive information. At the end of the study, we recruited $N = 188$ advisees and sent them the advisors' recommendation in this task.

Explanation of Advisors' Choices. In the second wave of data collection of the ChoiceFree experiment, we added an open-ended question asking participants to explain how they made their decision about order of information. 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?”. Two independent raters, who were blind to advisors' choices, coded the responses of advisors from the AMT-2 wave of the experiment and the 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 Online Appendix C.2.4.

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

B.4 Exclusion Criteria

In all of the experiments, participants who failed to answer the attention check correctly were not randomized into treatments and therefore, as pre-registered, they were excluded from completing the experiment. Further, we pre-registered that we would exclude participants who provide non-monotone responses to the multiple price list used to measure selfishness. Nonmonotone participants are $N = 28$ (8.6%), in the NoChoice experiment, $N = 676$ (10.8%) across the three waves of the Choice experiment, $N = 209$ (12.43%) in ChoiceStakes experiment, and $N = 51$ (9.3%) in the Information Architect experiment. For the analyses in the main text, we apply this exclusion criteria across studies with the exception of the ChoiceFree-Professionals sample for whom we did not collect this measure. In Online Appendix C.4, we repeat the main analyses including these participants. On top of this, all of our studies systematically exclude duplicate responses and participants classified as bots by Qualtrics bot detection feature.¹ Finally, since in all regressions we control for gender and age, participants with missing information for these variables are dropped from the analyses ($N = 11$ in the Choice Experiment, $N = 1$ in the ChoiceStakes Experiment; we did not have missing demographics in the other experiments). In the Prediction experiment, as an extra measure of attention, we ask participants to give a reason for their predictions by writing one sentence. As pre-registered, we exclude participants who provide answers to this question that are unrelated to the experiment, as determined by a research assistant.

¹In particular, we exclude participants with a Q_RecaptchaScore score lower than 0.5 on a scale from 0-1, which indicates a high probability that a given response comes from a bot, see <https://www.qualtrics.com/support/survey-platform/survey-module/survey-checker/fraud-detection/#BotDetection>.

C Additional Analyses

C.1 NoChoice Experiment

The tables below show regression analyses for advisors in the NoChoice treatment. As pre-registered, Table C.1 focuses on advisors who gave consistent answers in the elicitation of selfishness. Table C.2 includes all advisors.

Table C.1: Recommendations

	(1) Conflict	(2) No Conflict	(3) Both
See Incentive First	0.142** (0.062)	0.030 (0.078)	0.148** (0.061)
No Conflict			0.260*** (0.074)
See Incentive First X No Conflict			-0.130 (0.098)
Incentive for B	-0.193*** (0.062)	-0.066 (0.081)	-0.175*** (0.050)
Selfishness	0.108*** (0.028)	-0.026 (0.035)	0.076*** (0.023)
Constant	0.696*** (0.112)	0.935*** (0.171)	0.696*** (0.097)
Observations	213	86	299
R^2	0.137	0.028	0.124

Note: This table displays the estimated coefficients from linear probability models on the advisors' recommendations. See Incentive first is a binary indicator coded as 1 for participants who were randomly assigned to see the incentive first. Selfishness is standardized measure of the number of times the advisor chose to recommend the incentivized product in the measure of moral costs. The sample includes attentive participants who did not switch multiple times in the elicitation of selfishness. The regression includes individual controls for the advisor's gender and age. Robust standard errors (HC3) in parentheses.

Table C.2: Recommendations including Inattentive

	(1) Conflict	(2) No Conflict	(3) Both
See Incentive First	0.172*** (0.059)	0.096 (0.083)	0.173*** (0.059)
No Conflict			0.206*** (0.076)
See Incentive First X No Conflict			-0.082 (0.099)
Incentive for B	-0.217*** (0.062)	-0.089 (0.085)	-0.181*** (0.050)
Constant	0.735*** (0.108)	0.827*** (0.180)	0.707*** (0.095)
Observations	232	95	327
R^2	0.091	0.030	0.088

Note: This table displays the estimated coefficients from linear probability models on the advisors' recommendations. See Incentive first is a binary indicator coded as 1 for participants who were randomly assigned to see the incentive first. The sample includes all participants, including those who switched multiple times in the measure of morality. The regression includes individual controls for the advisor's gender and age. Robust standard errors (HC3) in parentheses.

C.2 Choice experiment

C.2.1 Preferences

Table C.3 breaks down preferences for information order by treatment and wave of the Choice Experiment. In Tables C.4 and C.5 we report the correlation between preferences to see the incentive first and preferences for blinding.

Table C.3: Preferences for Information Order by Wave

(1)	(2)	(3)	(4)
Wave	Treatment	N	Demand to See Incentive First
AMT-1	Choice Free	1308	0.563
AMT-1	Incentive First Costly	1347	0.422
AMT-2	Choice Free (\$0.15 commission)	511	0.628
AMT-2	Choice Free (\$15 for 1/100 commission)	542	0.565
AMT-3	Choice Free	213	0.451
AMT-3	Choice Free - Highx10	275	0.556
AMT-3	Choice Free - Highx100	110	0.600
AMT-3	Quality First Costly	1067	0.619
AMT-3	Incentive First Costly	215	0.340
Professionals	ChoiceFree—Professionals	712	0.480

Table C.4: Preferences for Blindness and Preferences for Information Order

	(1)	(2)	(3)
	Advisor Preference to Blind		
	Assigned Pref.	Not Assigned Pref.	Both
Prefer Incentive First	-0.217*** (0.030)	-0.268*** (0.051)	-0.216*** (0.030)
No Conflict	-0.044 (0.032)	-0.092* (0.053)	-0.057** (0.027)
Not Assigned Preference			0.010 (0.045)
Prefer Incentive First X Not Assigned Preference			-0.054 (0.058)
See Incentive First Costly	0.039 (0.057)	-0.101 (0.095)	-0.003 (0.049)
Assess Quality First Costly	0.000 (0.043)	-0.070 (0.074)	-0.022 (0.037)
Constant	0.550*** (0.064)	0.517*** (0.109)	0.545*** (0.056)
Observations	1121	363	1484
R^2	0.053	0.104	0.060

Notes: This table displays the coefficient estimates of OLS regressions on the advisor's preferences to blind themselves to incentives information in the Blinding task. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

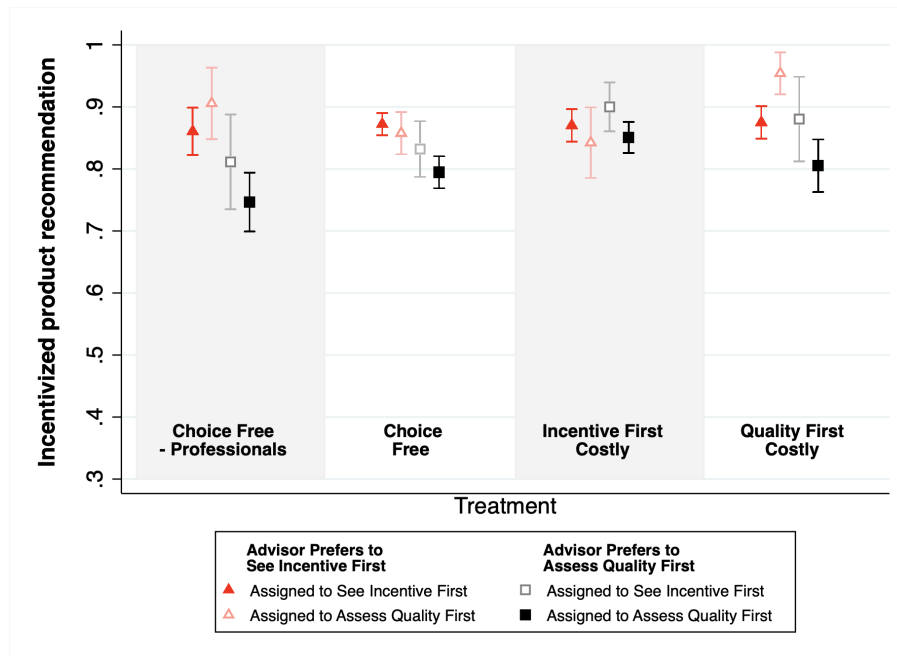
Table C.5: Preferences for Blindness, Information Order & Selfishness

	(1)	(2)	(3)
	Advisor Preference to Blind		
	Assigned Pref.	Not Assigned Pref.	Both
Prefer Incentive First	-0.215*** (0.030)	-0.260*** (0.052)	-0.213*** (0.030)
No Conflict	-0.046 (0.032)	-0.100* (0.053)	-0.060** (0.027)
Not Assigned Preference			0.008 (0.046)
Prefer Incentive First X Not Assigned Preference			-0.050 (0.058)
See Incentive First Costly	0.046 (0.057)	-0.110 (0.094)	-0.001 (0.048)
Assess Quality First Costly	0.006 (0.043)	-0.086 (0.073)	-0.023 (0.037)
Selfishness	-0.055*** (0.015)	-0.050** (0.025)	-0.052*** (0.013)
Constant	0.555*** (0.064)	0.536*** (0.108)	0.554*** (0.056)
Observations	1121	363	1484
R^2	0.064	0.114	0.070

Notes: This table displays the coefficient estimates of OLS regressions on the advisor's preferences to blind themselves to incentives information in the Blinding task, controlling for selfishness. Selfishness was elicited at the end of the experiment, using a multiple price list (MPL) with 5 decisions. The variable is a standardized measure of the number of times the advisor chose to recommend the incentivized product in the MPL task. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C.2.2 Recommendations

Figure C.1 displays recommendations for the cases in which advisors did not face a conflict of interest.



Notes: This figure presents the covariate-adjusted recommendations of the incentivized product for cases in which there was no conflict of interest. Error bars indicate ± 1 SE.

Figure C.1: Advisor Recommendations - No Conflict

Table C.6 reports the regression results for recommendations and looks at the relationship between recommendations and selfishness. In Tables C.7 and C.8, we break down the results for recommendations by product.

Table C.6: Advisor Recommendations

<i>Assignment:</i>	(1)	(2)	(3)
	Recommend incentivized product		
	Assigned Pref.	Not Assigned Pref.	Both
Prefer to See Incentive First	0.179*** (0.017)	-0.001 (0.031)	0.166*** (0.016)
Not Assigned Preference			0.055** (0.023)
Prefer to See Incentive First X Not Assigned Pref.			-0.132*** (0.027)
No Conflict	0.259*** (0.021)	0.210*** (0.036)	0.242*** (0.020)
No Conflict X Prefer to See Incentive First	-0.133*** (0.027)	0.002 (0.048)	-0.099*** (0.023)
No Conflict X Not Assigned Preference			0.020 (0.027)
See Incentive First Costly	0.033* (0.017)	0.018 (0.032)	0.029* (0.015)
Assess Quality First Costly	0.003 (0.029)	0.101* (0.052)	0.030 (0.026)
Incentive for B	-0.155*** (0.013)	-0.177*** (0.024)	-0.161*** (0.012)
Selfishness	0.054*** (0.006)	0.036*** (0.011)	0.049*** (0.005)
Female	0.001 (0.013)	-0.013 (0.024)	-0.002 (0.012)
Age	-0.002*** (0.001)	-0.002** (0.001)	-0.002*** (0.000)
Constant	0.731*** (0.029)	0.851*** (0.050)	0.747*** (0.026)
Observations	3915	1281	5196
R^2	0.114	0.086	0.104

Notes: This table displays the estimated coefficients from linear probability models on the advisor's decision to recommend the incentivized option. Column (1) focuses on individuals who are assigned their preference, while column (2) focuses on individuals who are not assigned their preference. Both groups are merged in column (3). Prefer to See Incentive First is an indicator of the advisor's preference, and Not Assigned Preference is an indicator for not receiving the preferred order. No Conflict is an indicator for the cases in which the signal of quality is not in conflict with the advisor's commission. Choice Free-Professionals, See Incentive First Costly and Assess Quality First Costly are indicator variables that take value 1 in the respective treatment, 0 otherwise. Selfishness was elicited at the end of the experiment, using a multiple price list (MPL) with 5 decisions. The variable is a standardized measure of the number of times the advisor chose to recommend the incentivized product in the MPL task. All regression models include controls for each wave of the experiment, whether incentives were probabilistic, the position of the products on the screen and the interaction between these two variables. Robust standard errors (HC3) in parentheses. * p<.10; ** p<.05; *** p<.01

Table C.7: Advisor Recommendations: Incentive for A

<i>Assignment:</i>	(1)	(2)	(3)
	Recommend incentivized product		
	Assigned Pref.	Not Assigned Pref.	Both
Prefer to See Incentive First	0.180*** (0.020)	-0.008 (0.035)	0.172*** (0.019)
Not Assigned Preference			0.073*** (0.027)
Prefer to See Incentive First X Not Assigned Pref.			-0.161*** (0.033)
No Conflict	0.202*** (0.027)	0.170*** (0.038)	0.182*** (0.025)
No Conflict X Prefer to See Incentive First	-0.112*** (0.032)	0.035 (0.050)	-0.077*** (0.027)
No Conflict X Not Assigned Preference			0.048* (0.029)
See Incentive First Costly	0.022 (0.023)	0.044 (0.041)	0.027 (0.020)
Assess Quality First Costly	0.002 (0.039)	-0.008 (0.066)	-0.002 (0.033)
Female	0.017 (0.016)	-0.001 (0.030)	0.013 (0.014)
Age	-0.002** (0.001)	-0.001 (0.001)	-0.002*** (0.001)
Constant	0.735*** (0.035)	0.785*** (0.061)	0.730*** (0.032)
Observations	2242	725	2967
R^2	0.074	0.048	0.065

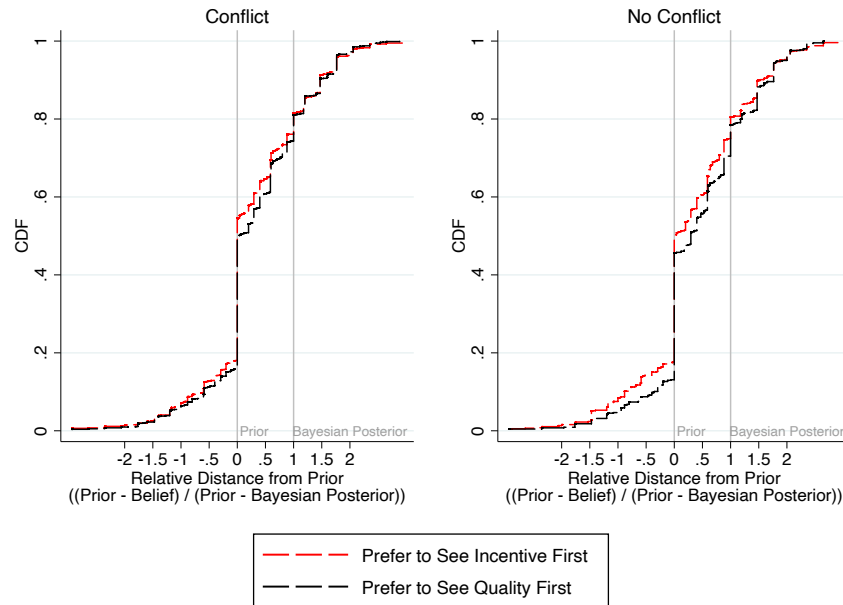
Notes: This table displays the estimated coefficients from linear probability models on the advisor's decision to recommend the incentivized option. Column (1) focuses on individuals who are assigned their preference, while column (2) focuses on individuals who are not assigned their preference. Both groups are merged in column (3). Prefer to See Incentive First is an indicator of the advisor's preference, and Not Assigned Preference is an indicator for not receiving the preferred order. No Conflict is an indicator for the cases in which the signal of quality is not in conflict with the advisor's commission. Choice Free-Professionals, See Incentive First Costly and Assess Quality First Costly are indicator variables that take value 1 in the respective treatment, 0 otherwise. All regression models include controls for each wave of the experiment, whether incentives were probabilistic, the position of the products on the screen and the interaction between these two variables. Robust standard errors (HC3) in parentheses. * p<.10; ** p<.05; *** p<.01

Table C.8: Advisor Recommendations: Incentive for B

<i>Assignment:</i>	(1)	(2)	(3)
	Recommend incentivized product		
	Assigned Pref.	Not Assigned Pref.	Both
Prefer to See Incentive First	0.216*** (0.027)	0.018 (0.048)	0.197*** (0.025)
Not Assigned Preference			0.047 (0.034)
Prefer to See Incentive First X Not Assigned Pref.			-0.122*** (0.039)
No Conflict	0.295*** (0.028)	0.229*** (0.049)	0.275*** (0.026)
No Conflict X Prefer to See Incentive First	-0.157*** (0.038)	-0.011 (0.067)	-0.120*** (0.033)
No Conflict X Not Assigned Preference			0.008 (0.038)
See Incentive First Costly	0.049* (0.027)	-0.000 (0.047)	0.035 (0.023)
Assess Quality First Costly	0.003 (0.045)	0.201** (0.082)	0.055 (0.040)
Female	-0.009 (0.019)	-0.024 (0.034)	-0.015 (0.017)
Age	-0.002*** (0.001)	-0.004*** (0.001)	-0.003*** (0.001)
Constant	0.562*** (0.041)	0.740*** (0.074)	0.599*** (0.037)
Observations	2206	735	2941
R^2	0.097	0.087	0.088

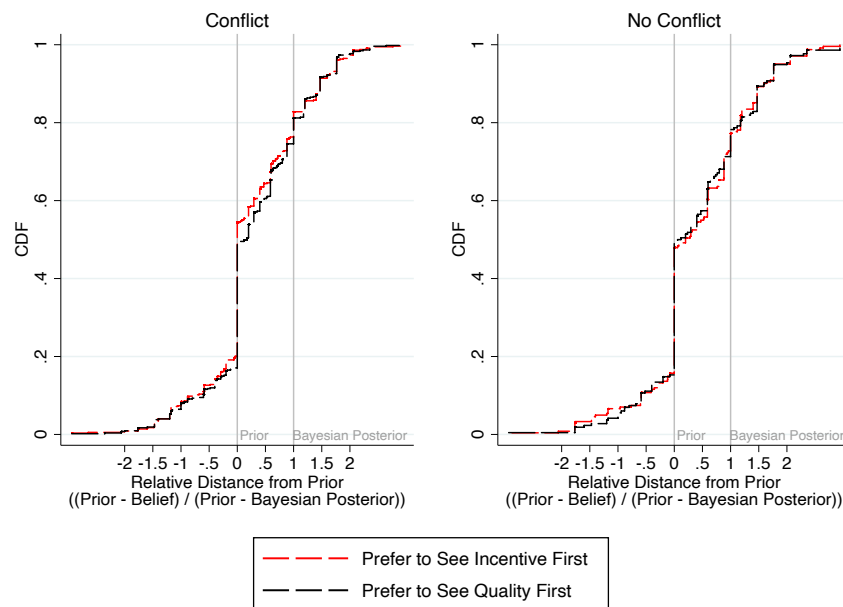
Notes: This table displays the estimated coefficients from linear probability models on the advisor's decision to recommend the incentivized option. Column (1) focuses on individuals who are assigned their preference, while column (2) focuses on individuals who are not assigned their preference. Both groups are merged in column (3). Prefer to See Incentive First is an indicator of the advisor's preference, and Not Assigned Preference is an indicator for not receiving the preferred order. No Conflict is an indicator for the cases in which the signal of quality is not in conflict with the advisor's commission. Choice Free-Professionals, See Incentive First Costly and Assess Quality First Costly are indicator variables that take value 1 in the respective treatment, 0 otherwise. All regression models include controls for each wave of the experiment, whether incentives were probabilistic, the position of the products on the screen and the interaction between these two variables. Robust standard errors (HC3) in parentheses. * p<.10; ** p<.05; *** p<.01

C.2.3 Beliefs



Notes: This figure presents the cumulative distribution of the extent of belief updating by advisors. This measure is the ratio of the difference between the advisor's belief and the prior, divided by the difference between the Bayesian posterior and the prior. The figure focuses on cases in which advisors are assigned to their preferred information order.

Figure C.2: Beliefs - Assigned Preferred Information Order



Notes: This figure presents the cumulative distribution of the extent of belief updating by advisors. This measure is the ratio of the difference between the advisor's belief and the prior, divided by the difference between the Bayesian posterior and the prior. The figure focuses on cases in which advisors are not assigned to their preferred information order.

Figure C.3: Beliefs - Not Assigned Preferred Information Order

Table C.9: Belief Updating when Signal is \$0

	(1)	(2)	(3)	(4)
	Log-odds Belief			
<i>Assignment:</i>	Assigned Pref.	Not Assigned Pref.	Assigned Pref.	Not Assigned Pref.
<i>Data:</i>		All	Excl. update in wrong direction	
Panel A: Pooled				
β_C	0.217*** (0.020)	0.259*** (0.036)	0.449*** (0.017)	0.506*** (0.031)
β_{NC}	0.309*** (0.038)	0.287*** (0.064)	0.553*** (0.032)	0.538*** (0.052)
Panel B: By Choice of Information Order				
$\beta_C^{f=i}$	0.173*** (0.028)	0.265*** (0.049)	0.419*** (0.024)	0.517*** (0.042)
$\beta_C^{f=q}$	0.262*** (0.029)	0.253*** (0.052)	0.479*** (0.024)	0.495*** (0.045)
$\beta_{NC}^{f=i}$	0.285*** (0.052)	0.361*** (0.086)	0.552*** (0.045)	0.562*** (0.069)
$\beta_{NC}^{f=q}$	0.339*** (0.056)	0.194** (0.096)	0.555*** (0.046)	0.503*** (0.080)
Observations	1765	569	1428	453
$\beta_C^{f=q} = \beta_C^{f=i}$	0.026	0.872	0.076	0.727
$\beta_{NC}^{f=q} = \beta_{NC}^{f=i}$	0.476	0.193	0.964	0.576

Notes: The outcome in all regressions is the log belief ratio, when the advisors sees a \$0 ball for product B. β_C^f and β_{NC}^f are the estimated effects of the log likelihood ratio for conflict and no conflict signals, respectively, for advisors who prefer order ($f = i$ indicates a preference to see the incentive first, and $f = q$ indicates a preference to see quality first). Columns(1) and (2) include all advisors. Columns (3) and (4) exclude advisors who updated in the wrong direction. Columns (1) and (3) include only advisors who were assigned their preference, while columns (2) and (4) include only advisors who were not assigned their preference. Robust standard errors (HC3) in parentheses.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.10: Belief Updating when Signal is \$2

	(1)	(2)	(3)	(4)
	Log-odds Belief			
<i>Assignment:</i>	Assigned Pref.	Not Assigned Pref.	Assigned Pref.	Not Assigned Pref.
<i>Data:</i>		All	Excl. update in wrong direction	
Panel A: Pooled				
β_C	0.471*** (0.026)	0.410*** (0.043)	0.726*** (0.023)	0.506*** (0.031)
β_{NC}	0.473*** (0.039)	0.484*** (0.066)	0.757*** (0.034)	0.538*** (0.052)
Panel B: By Choice of Information Order				
$\beta_C^{f=i}$	0.436*** (0.035)	0.363*** (0.060)	0.702*** (0.031)	0.661*** (0.053)
$\beta_C^{f=q}$	0.511*** (0.038)	0.460*** (0.063)	0.752*** (0.034)	0.732*** (0.055)
$\beta_{NC}^{f=i}$	0.380*** (0.058)	0.460*** (0.106)	0.724*** (0.048)	0.826*** (0.085)
$\beta_{NC}^{f=q}$	0.573*** (0.053)	0.508*** (0.080)	0.789*** (0.048)	0.700*** (0.072)
Observations	2620	878	2246	740
$\beta_C^{f=q} = \beta_C^{f=i}$	0.149	0.265	0.276	0.352
$\beta_{NC}^{f=q} = \beta_{NC}^{f=i}$	0.014	0.715	0.339	0.257

Notes: The outcome in all regressions is the log belief ratio, when the advisors sees a \$2 ball for product B. β_C^f and β_{NC}^f are the estimated effects of the log likelihood ratio for conflict and no conflict signals, respectively, for advisors who prefer order ($f = i$ indicates a preference to see the incentive first, and $f = q$ indicates a preference to see quality first). Columns(1) and (2) include all advisors. Columns (3) and (4) exclude advisors who updated in the wrong direction. Columns (1) and (3) include only advisors who were assigned their preference, while columns (2) and (4) include only advisors who were not assigned their preference. Robust standard errors (HC3) in parentheses.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.11: Belief Updating: Correct Choice

	(1)	(2)	(3)	(4)
	Belief Correct			
<i>Assignment:</i> <i>Data:</i>	Assigned Pref.	Not Assigned Pref. All	Assigned Pref. Excl. update in wrong direction	Not Assigned Pref.
Prefer to See Incentive First	-0.039*** (0.013)	-0.032 (0.024)	-0.042*** (0.016)	-0.037 (0.028)
No Conflict	0.018 (0.019)	0.012 (0.033)	0.015 (0.022)	0.005 (0.038)
No Conflict X Prefer to See Incentive First	0.012 (0.025)	0.022 (0.045)	0.019 (0.029)	0.037 (0.052)
See Incentive First Costly	-0.002 (0.015)	-0.003 (0.027)	-0.001 (0.018)	-0.007 (0.032)
Assess Quality First Costly	0.036 (0.025)	-0.046 (0.047)	0.042 (0.030)	-0.051 (0.052)
Incentive for B	-0.001 (0.011)	0.014 (0.021)	-0.002 (0.013)	0.017 (0.024)
Female	-0.029** (0.011)	-0.016 (0.020)	-0.033** (0.013)	-0.011 (0.024)
Age	-0.000 (0.000)	-0.001 (0.001)	0.000 (0.001)	-0.001 (0.001)
Constant	0.196*** (0.024)	0.192*** (0.044)	0.221*** (0.028)	0.218*** (0.050)
Observations	4448	1460	3700	1224
R^2	0.007	0.009	0.008	0.010

Notes: This table displays the estimated coefficients from linear probability models on the advisor's beliefs that the quality of product B is low measured via their choice of one out of 10 possible belief bins (ranging from 0 to 100, in steps of 10). Column (1) focuses on individuals who are assigned their preference, while column (2) focuses on individuals who are not assigned their preference. Columns (3) and (4) exclude individuals who chose a bin that is consistent with updating in the incorrect direction. Prefer to See Incentive First is an indicator of the advisor's preference, and Not Assigned Preference is an indicator for not receiving the preferred order. No Conflict is an indicator for the cases in which the signal of quality is not in conflict with the advisor's commission. Choice Free-Professionals, See Incentive First Costly and Assess Quality First Costly are indicator variables that take value 1 in the respective treatment, 0 otherwise. All regression models include controls for each wave of the experiment, whether incentives were probabilistic, the position of the products on the screen and the interaction between these two variables. Robust standard errors (HC3) in parentheses. * $p < .10$; ** $p < .05$; *** $p < .01$

C.2.4 Explanations for Choices

In the second wave of the Choice experiment (AMT-2) and the Choice Free Professional treatment of the Choice Experiment, we asked advisors to explain their choices of information order. A total of $N = 1,747$ advisors ($N = 1,035$ from AMT-2 and $N = 712$ professionals) from the main sample answered this question. We then had two independent raters code the reasons for choosing to see the incentive first or quality first. 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.

Table C.12 below reports the raters' coding of the 91% of the data ($N = 1554$ advisors, of which $N = 660$ were professionals) of advisors who provided an explanation.

Table C.12: Advisors' Explanations: Detailed Results

		Advisors' Explanations of Preference (Categories)			
		Limiting Bias	Indifference	Commission	Other reasons
Sample: All (N=1,554)					
Prefer to:	Assess Quality First	47.0%	10.9%	7.8%	37.4%
	See Incentive First	5.9%	7.1%	36.4%	55.1%
Sample: AMT (N=894)					
Prefer to:	Assess Quality First	41.3%	11.4%	10.7%	39.4%
	See Incentive First	5.1%	7.5%	36.4%	55.1%
Sample: Professionals (N=660)					
Prefer to:	Assess Quality First	53.1%	10%	4.9%	35.4%
	See Incentive First	7.2%	6.6%	36.4%	55.2%

Note: This table displays the average rating 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.

C.3 Comparing the Choice and NoChoice Experiments

In this section, we compare the recommendation decisions in the Choice experiment to those in the NoChoice experiment. For this comparison, we focus on the ChoiceFree treatment, which uses the same sample and incentives of the NoChoice experiment, and cases in which the advisor's incentives were in conflict with the quality signal.

When advisors are assigned to see the Incentive first in the NoChoice experiment, we estimate a 16.9pp increase in recommendations as compared to advisors who are randomly assigned to see Quality first (Table C.13 column (1)). Instead, when advisors prefer and are assigned to see the incentive first, we estimated a 23.5pp increase in recommendations, as compared to advisors who prefer and are assigned to see quality first (Table C.13 column (2)). This increase is 7.6 percentage points larger, albeit not

statistically different, than the increase in recommendations observed when advisors see the incentive first in NoChoice.

In addition to examining the difference in recommendations between advisors who choose (Choice experiment) or are assigned (NoChoice Experiment) an information order, we also conduct exploratory comparisons of the levels. Table C.14 reports covariate-adjusted recommendation rates for the See Incentive First and See Quality first treatments of the NoChoice experiment, and for assignment to see the Incentive first (vs. see Quality First) conditional on preferences for the Choice experiment.

We first focus on advisors who choose or are assigned to seeing the incentive first. In the NoChoice experiment, advisors who are assigned to “See Incentive First” recommend the incentivized option in 79% of the cases. In the Choice experiment, advisors who prefer to see the incentive first and are assigned to see the incentive first recommend the incentivized option in 81% of the cases. When advisors prefer to see quality first but are assigned to see the incentive first, they instead recommend the incentivized option in 67.9% of the cases. This difference in recommendations is in line with advisors who prefer to see the incentive first being more selfish. In the experiment, 55% of advisors prefer to see the incentive first whereas 45% of advisors prefer to see quality first. Weighting by their preference, we can estimate what the recommendation rate would have been if advisors in the Choice experiment were not asked to make a choice, and were rather randomly assigned to see the incentive first as in the NoChoice Experiment. In that case, the predicted rate of recommendations is 76.6% (from $0.55 \times 0.810 + 0.45 \times 0.679$). This recommendation rate is 2.4pp smaller than the rate of 79% we observe in the NoChoice experiment, but falls within its 95% confidence interval. This small decrease in recommendation could be due to some advisors being limited in their ability to justify self-serving recommendations or due to noise in the data.

Next, we consider the preference to see quality first. In the NoChoice experiment, advisors assigned to this preference order recommend the incentivized option in 62% of the cases. In the Choice experiment, advisors who prefer and are assigned to see quality first recommend the incentivized option in 56.9% of the cases, while advisors who prefer to see the incentive first but are assigned to see quality first recommend the incentivized option in 71.3% of the cases. Weighting by advisors’ preferences, we estimate that the recommendation rate if advisors had been randomly assigned to see quality first as in the NoChoice experiment would have been 63%. This estimate is very close to the 62% of recommendations we estimate in the NoChoice experiment.

Taken together, the results observed in the Choice experiment are consistent with those observed in the NoChoice experiment. Advisor recommendations in the Choice experiment are still significantly affected by assignment to the advisors’ preferred order,

which indicates that their active choice did not remove the scope for self-deception. At the same time, the 2.4pp difference between our prediction from the Choice data and the results of the NoChoice experiment in our exploratory analyses, might potentially indicate that the scope for self-deception may have been directionally restricted.

Table C.13: Advisor Recommendations

	(1)	(2)	(3)
	Recommend incentivized product		
<i>Sample:</i>	NoChoice.	Choice	Both
<i>Assignment:</i>	Assigned	Prefer and Assigned	Both
See Incentive First	0.1686***	0.2352***	0.2359***
No Choice	(0.0575)	(0.0226)	(0.0226)
See IncentiveFirst X NoChoice			0.0471
No Conflict			(0.0444)
No Choice X No Conflict			-0.0764
See Incentive First X No Conflict			(0.0605)
See Incentive First X No Choice X No Conflict			0.0006
Incentive for B			(0.0809)
Constant	-0.1418	-0.1748***	-0.1752***
Observations	(0.1063)	(0.0411)	(0.0412)
R^2			0.0361
			(0.1125)
	-0.1634***	-0.1514***	-0.1527***
	(0.0495)	(0.0191)	(0.0178)
	0.6918***	0.7456***	0.7317***
	(0.0921)	(0.0371)	(0.0353)
	299	1931	2230
	0.093	0.117	0.113

Note: This table displays the estimated coefficients from linear probability models on the advisor's decision to recommend the incentivized option. Column (1) focuses on the NoChoice Experiment, while column (2) focuses on the Choice Experiment (ChoiceFree Treatment only) and on individuals who are assigned their preference. Both groups are merged in column (3). See Incentive First is an indicator for whether advisors are randomly assigned to see the incentive first in NoChoice, and whether, conditional on preferring to see the incentive first, they are assigned to see the incentive first in Choice. No Conflict is an indicator for the cases in which the signal of quality is not in conflict with the advisor's commission. All regression models include individual controls for the advisor's gender and age, each wave of the experiment, whether incentives were probabilistic, the position of the products on the screen and the interaction between these two variables. Robust standard errors (HC3) in parentheses. * p<.10; ** p<.05; *** p<.01

Table C.14: Recommendations in the NoChoice and Choice Experiment

Experiment	Treatment	Mean	95% CI
NoChoice	See Incentive First	79%	[71%-87%]
	See Quality First	62%	[53%-71%]
Choice	Prefer to See Incentive First & Assigned to See Incentive First	81%	[73%-84%]
	Prefer to See Incentive First & Assigned to See Quality First	71%	[66%-77%]
	Prefer to See Quality First & Assigned to See Incentive First	68%	[61%-74%]
	Prefer to See Quality First & Assigned to See Quality First	57%	[53%-61%]
Choice - Predicted	See Incentive First	77%	-
	See Quality First	63%	-

Note: This table displays covariate-adjusted estimates of frequency of recommendations of the incentivized product by treatment and assignment in the Choice and NoChoice experiment, obtained via OLS regression.

C.4 The Higher Incentives Treatments

In this section, we report the results from the two robustness treatments that we collected as part of the AMT-3 wave of the choice experiment. In these treatments, we scale the incentives by a factor of 10 (High Stakes - 10-fold incentives) or 100 (High Stakes - 100-fold incentives). As part of that wave, we also collected data for our regular version of the Choice Free treatment with low incentives (a \$0.15 commission and products that yielded \$0 or \$2 to the client). As shown in Table C.3, the share of advisors who choose to see the incentive first is larger when advisors face larger incentives (45% with regular incentives as opposed to 55% with 10-fold incentives and 60% with 100-fold incentives; $p = 0.02$ and $p = 0.01$, respectively). This data shows that despite the substantially higher incentives, the fraction of advisors who prefers to assess quality first remains substantial.

As shown in Table C.15, when looking at preferences for information order in our full sample using OLS regressions and controlling for wave, we see that advisors in the treatments with higher incentives are 9 (High Stakes - 10 fold incentives) and 13 (High Stakes - 100 fold incentives) percentage points more likely to choose to see the incentives first than participants who were presented with smaller incentives.

In Table C.16, we report the results for recommendations. As displayed in the table, advisors in the these treatments are directionally more likely to recommend the incentivized product than those who faced smaller incentives in the Choice Free treatment. Importantly, the coefficient for the interaction between preferring to see the incentive first and the indicator for these treatments is not statistically significant (directionally, it is positive). Taken together, these results suggest that the effect of information order is robust to increasing the stakes.

Table C.15: Preference for Information Order: Including Incentives Treatments

	(1)	(2)	(3)
	Prefer to See Incentive First		
See Incentive First Costly	-0.14*** (0.02)	-0.14*** (0.02)	-0.14*** (0.02)
Assess Quality First Costly	0.15*** (0.03)	0.15*** (0.03)	0.15*** (0.03)
Choice Free – Professionals	-0.10*** (0.03)		
High Stakes (10-fold incentives)	0.09** (0.04)	0.09** (0.04)	0.09** (0.04)
High Stakes (100-fold incentives)	0.13** (0.05)	0.13** (0.05)	0.13** (0.05)
Selfishness		0.03*** (0.01)	0.04*** (0.01)
See Incentive First Costly X Selfishness			-0.02 (0.02)
See Quality First Costly X Selfishness			-0.02 (0.02)
Female	-0.03** (0.01)	-0.02 (0.01)	-0.02 (0.01)
Age	-0.00*** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)
Constant	0.68*** (0.02)	0.67*** (0.02)	0.67*** (0.02)
Observations	6293	5581	5581
R^2	0.034	0.039	0.039

Notes: This table displays the estimated coefficients from linear probability models on the preference to see the incentive first. See Incentive First Costly and Assess Quality First Costly are indicator variables that take value 1 in the respective treatment, 0 otherwise. Selfishness was elicited at the end of the experiment, using a multiple price list (MPL) with 5 decisions. The variable is a standardized measure of the number of times the advisor chose to recommend the incentivized product in the MPL task. The regression models in columns (2) and (3) include individual controls for the advisor's gender and age, each wave of the experiment, whether incentives were probabilistic, the position of the products on the screen and the interaction between these two variables. Robust standard errors (HC3) in parentheses. * $p < .10$; ** $p < .05$; *** $p < .01$

Table C.16: Advisor Recommendations: Including Incentives Treatments

<i>Assignment:</i>	(1)	(2)	(3)
	Recommend incentivized product		
	Assigned Pref.	Not Assigned Pref.	Both
Prefer to See Incentive First	0.196*** (0.016)	0.002 (0.029)	0.182*** (0.015)
Not Assigned Preference			0.061*** (0.021)
Prefer to See Incentive First X Not Assigned Pref.			-0.140*** (0.026)
No Conflict	0.256*** (0.020)	0.201*** (0.033)	0.236*** (0.018)
No Conflict X Prefer to See Incentive First	-0.137*** (0.025)	0.013 (0.045)	-0.098*** (0.022)
No Conflict X Not Assigned Preference			0.019 (0.025)
See Incentive First Costly	0.035** (0.017)	0.020 (0.032)	0.031** (0.015)
Assess Quality First Costly	0.004 (0.030)	0.091* (0.052)	0.027 (0.026)
Incentive for B	-0.168*** (0.012)	-0.182*** (0.022)	-0.171*** (0.011)
Female	0.006 (0.012)	-0.031 (0.022)	-0.004 (0.011)
Age	-0.002*** (0.001)	-0.003*** (0.001)	-0.002*** (0.000)
High Stakes (10-fold incentives)	0.119* (0.064)	0.133 (0.100)	0.133** (0.061)
High Stakes (100-fold incentives)	0.131 (0.096)	0.291 (0.180)	0.142 (0.093)
Prefer to See Incentive First X High Stakes (10-fold)	0.038 (0.067)	0.013 (0.131)	0.041 (0.064)
Prefer to See Incentive First X High Stakes (100-fold)	0.044 (0.104)	-0.403 (0.248)	0.052 (0.101)
Constant	0.734*** (0.027)	0.867*** (0.047)	0.753*** (0.024)
Observations	4743	1550	6293
R^2	0.110	0.085	0.101

Notes: This table displays the estimated coefficients from linear probability models on the advisor's decision to recommend the incentivized option. Column (1) focuses on individuals who are assigned their preference, while column (2) focuses on individuals who are not assigned their preference. Both groups are merged in column (3). Prefer to See Incentive First is an indicator of the advisor's preference, and Not Assigned Preference is an indicator for not receiving the preferred order. No Conflict is an indicator for the cases in which the signal of quality is not in conflict with the advisor's commission. Choice Free-Professionals, See Incentive First Costly and Assess Quality First Costly are indicator variables that take value 1 in the respective treatment, 0 otherwise. All regression models include controls for each wave of the experiment, whether incentives were probabilistic, the position of the products on the screen and the interaction between these two variables. Robust standard errors (HC3) in parentheses. * p<.10; ** p<.05; *** p<.01

C.5 Including Inattentive Participants

Our main sample exclude all participants who give non-monotone responses to the multiple price list questions that aims to measure selfishness and classify participants into moral types. A total of 1355 participants switched multiple times in the MPL, and, as pre-registered were therefore excluded from the main analyses. Here, we repeat the analyses for preferences and recommendations from the main text (Tables 2 and 3) but include participants who switch multiple times in the multiple price list to measure selfishness. Column 1 includes only attentive participants from the main sample, and Column 2 includes all participants (including the inattentive ones).

Table C.17: Preference for Information Order—Including Inattentive

	(1)	(2)
	Prefer to See Incentive First	
	Main Sample	Including Inattentive
See Incentive First Costly	-0.139*** (0.018)	-0.136*** (0.017)
Assess Quality First Costly	0.152*** (0.029)	0.154*** (0.028)
Choice Free – Professionals		-0.103*** (0.026)
Female	-0.026* (0.014)	-0.035*** (0.012)
Age	-0.002*** (0.001)	-0.003*** (0.001)
Constant	0.668*** (0.025)	0.683*** (0.023)
Observations	5196	6547
R^2	0.037	0.036

Note: This table displays the estimated coefficients from linear probability models on the advisor's preference to see the incentive first. Column (1) focuses on individuals who are assigned their preference, while column (2) focuses on individuals who are not assigned their preference. Both groups are merged in column (3). Prefer to See Incentive First is an indicator of the advisor's preference, and Not Assigned Preference is an indicator for not receiving the preferred order. No Conflict is an indicator for the cases in which the signal of quality is not in conflict with the advisor's commission. See Incentive First Costly and Assess Quality First Costly are indicator variables that take value 1 in the respective treatment, 0 otherwise. All regression models include individual controls for the advisor's gender and age, each wave of the experiment, whether incentives were probabilistic, the position of the products on the screen and the interaction between these two variables. Robust standard errors (HC3) in parentheses. * $p < .10$; ** $p < .05$; *** $p < .01$

Table C.18: Advisor Recommendations—Including Inattentive

<i>Assignment:</i>	(1)	(2)	(3)
	Recommend incentivized product		
	Assigned Pref.	Not Assigned Pref.	Both
Prefer to See Incentive First	0.181*** (0.015)	-0.002 (0.027)	0.167*** (0.015)
Not Assigned Preference			0.051** (0.021)
Prefer to See Incentive First X Not Assigned Preference			-0.128*** (0.025)
No Conflict	0.254*** (0.019)	0.195*** (0.032)	0.234*** (0.018)
No Conflict X Prefer to See Incentive First	-0.133*** (0.024)	0.026 (0.043)	-0.092*** (0.021)
No Conflict X Not Assigned Preference			0.019 (0.024)
Choice Free–Professionals	-0.023 (0.025)	0.051 (0.043)	-0.003 (0.022)
See Incentive First Costly	0.028* (0.017)	0.012 (0.030)	0.023 (0.015)
Assess Quality First Costly	-0.001 (0.029)	0.081 (0.050)	0.021 (0.025)
Incentive for B	-0.166*** (0.012)	-0.190*** (0.022)	-0.173*** (0.011)
Female	0.008 (0.012)	-0.016 (0.022)	0.001 (0.011)
Age	-0.002*** (0.001)	-0.003*** (0.001)	-0.002*** (0.000)
Constant	0.728*** (0.026)	0.881*** (0.046)	0.755*** (0.024)
Observations	4920	1627	6547
R^2	0.095	0.085	0.089

Note: This table displays the estimated coefficients from linear probability models on the advisor's decision to recommend the incentivized option. Column (1) focuses on individuals who are assigned their preference, while column (2) focuses on individuals who are not assigned their preference. Both groups are merged in column (3). Prefer to See Incentive First is an indicator of the advisor's preference, and Not Assigned Preference is an indicator for not receiving the preferred order. No Conflict is an indicator for the cases in which the signal of quality is not in conflict with the advisor's commission. Choice Free-Professionals, See Incentive First Costly and Assess Quality First Costly are indicator variables that take value 1 in the respective treatment, 0 otherwise. All regression models include controls for each wave of the experiment, whether incentives were probabilistic, the position of the products on the screen and the interaction between these two variables. Robust standard errors (HC3) in parentheses. * p<.10; ** p<.05; *** p<.01

C.6 The ChoiceStakes Experiment: Additional Results

Table C.19: Preference for Information Order

	(1)	(2)
	Prefer to See Incentive First	
Low Incentive	-0.276*** (0.027)	-0.278*** (0.027)
High Incentive	0.029 (0.031)	0.028 (0.031)
Selfishness		0.022* (0.012)
Female	-0.039 (0.024)	-0.036 (0.024)
Age	-0.001 (0.001)	-0.001 (0.001)
Constant	0.477*** (0.044)	0.469*** (0.044)
Observations	1471	1471
R^2	0.088	0.091

Note: This table displays the estimated coefficients from linear probability models on the preference to see the incentive first. Low Incentive and High Incentive are indicator variables that take value 1 in the respective treatment, 0 otherwise. Selfishness was elicited at the end of the experiment, using a multiple price list (MPL) with 5 decisions. The variable is a standardized measure of the number of times the advisor chose to recommend the incentivized product in the MPL task. Robust standard errors (HC3) in parentheses. * $p < .10$; ** $p < .05$; *** $p < .01$

Table C.20: Advisor Recommendations

<i>Assignment:</i>	Recommend incentivized product		
	(1) Assigned Pref.	(2) Not Assigned Pref.	(3) Both
Prefer to See Incentive First	0.1363*** (0.0512)	0.0456 (0.1023)	0.1498*** (0.0471)
Not Assigned Preference			0.0887** (0.0402)
Prefer to See Incentive First X Not Assigned Preference			-0.1515*** (0.0582)
No Conflict	0.2842*** (0.0344)	0.2456*** (0.0603)	0.2836*** (0.0329)
No Conflict X Prefer to See Incentive First	-0.1307** (0.0558)	-0.1912* (0.1093)	-0.1383*** (0.0498)
No Conflict X Not Assigned Preference			-0.0484 (0.0555)
Low Incentive	-0.1451*** (0.0403)	-0.0561 (0.0719)	-0.1249*** (0.0351)
Low Incentive X Prefer to See Incentive First	0.0121 (0.0855)	0.0775 (0.1704)	0.0134 (0.0764)
High Incentive	0.0241 (0.0442)	0.0535 (0.0751)	0.0323 (0.0379)
High Incentive X Prefer to See Incentive First	0.0748 (0.0636)	0.0067 (0.1197)	0.0565 (0.0560)
Incentive for B	-0.1269*** (0.0272)	-0.0869* (0.0518)	-0.1186*** (0.0239)
Female	-0.0317 (0.0273)	-0.1215** (0.0499)	-0.0535** (0.0239)
Age	-0.0017 (0.0012)	-0.0004 (0.0021)	-0.0014 (0.0010)
Constant	0.7145*** (0.0555)	0.7399*** (0.1023)	0.6987*** (0.0497)
Observations	1104	367	1471
R^2	0.121	0.063	0.104

Notes: This table displays the estimated coefficients from linear probability models on the advisor's decision to recommend the incentivized option. Column (1) focuses on individuals who are assigned their preference, while column (2) focuses on individuals who are not assigned their preference. Both groups are merged in column (3). Prefer to See Incentive First is an indicator of the advisor's preference, and Not Assigned Preference is an indicator for not receiving the preferred order. No Conflict is an indicator for the cases in which the signal of quality is not in conflict with the advisor's commission. Low Incentive and High Incentive are indicator variables that take value 1 in the respective treatment, 0 otherwise. Robust standard errors (HC3) in parentheses. * $p < .10$; ** $p < .05$; *** $p < .01$

C.7 The Information Architect Experiment: Additional Results

In the Information Architect experiment, we investigate preferences for information order of a third party who determines how advisors receive information. The sample is comprised by 498 attentive participants. An additional 51 participants switched multiple times in the task that measured selfishness. As preregistered, these participants are dropped from the main analysis, but for robustness, we repeat the analysis including these participants. For the experiment, we then recruited 498 participants to play the role of advisors and matched 1 out of 10 advisors with a client. Table C.21 presents regression results comparing IA preferences in IA-Advisor, relative to IA-Client (omitted category), controlling for the IA’s gender and age.

Table C.21: IA Preferences by condition

	(1)	(2)
	DM Choice to See Incentive First	
<i>Sample:</i>	Main Sample	Including Inattentive
IA-Advisor	0.148*** (0.045)	0.142*** (0.042)
Constant	0.334*** (0.082)	0.353*** (0.078)
Observations	498	549
R^2	0.033	0.031

Notes: This table displays the coefficient estimates of OLS regressions on the Information Architect’s preferences to have the advisor see the incentive first for the main sample (Column 1) and the sample that includes inattentive participants who switched multiple times in the selfishness measure. IA-Advisor is an indicator for whether advisors have an incentive to receive information about their incentive first. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

As part of the experiment, we then recruited $N = 498$ advisors, presented with the information order selected by the Information Architect and had them make their recommendation to the client. For this purpose, we recruited $N = 50$ clients for the main task; of these 86% followed the recommendations. We also recruited an additional $N = 50$ advisees for the MPL task that measured advisors’ moral costs and matched them with 1 out of 10 Information Architects, and an additional $N = 50$ advisees for the same task and matched them with 1 out of 10 advisors. Of these, 86% and 80% of advisees followed the recommendation.

D The NoChoiceSimultaneous Experiment

D.1 Experimental Design

To test how advisors behave when they receive information about their own incentive and information about the quality of the product (i.e., the quality signal) simultaneously, we conducted an additional wave of the NoChoice experiment. The experiment replicates the design and procedures of the NoChoice experiment. On top of the See Incentive First and See Quality First treatments, this wave of data collection added an additional treatment (Simultaneous) where information about incentive and the quality signal were presented to participants on the same screen. Participants were assigned to the three treatments at a 1:1:3 ratio, as we planned to merge the data with those of the original NoChoice experiment for the analyses. By comparing the rates of recommendations of the incentivized product in the Simultaneous treatment to those in the See Incentive First and See Quality First treatment, we can investigate how receiving information simultaneously affects recommendation in case of a conflict of interest.

Procedures. The experiment was conducted on Amazon Mechanical Turk (AMT); the design and analyses plan were pre-registered on aspredicted.org (#79521 and #82164). Participants received \$1 payment for taking part in the experiment and for making a recommendation. On top of that, they received a \$.15 commission for recommending either Product A or Product B. Advisors were informed that one out of 10 advisors would be matched with a client, another AMT participant, and their advice was delivered to them.

For the See Incentive First and See Quality First treatments, all procedures were identical to those of the NoChoice Experiment, with some small modifications. In particular, to address potential concerns about demand effect, whereby participants may assume that the order of information is determined by the experimenter thereby leading participants to “react” to the experimenter decisions, we informed participants that the order of information in the experiment was randomly determined by the computer. Further, we also informed participants that whether the commission was for Product A or Product B was randomly determined by the computer. In the See Incentive First treatment, participants first saw information about what product yielded a commission and then received further information about the quality of the product. In the See Quality First treatment, participants first learned about the quality of the product and then received information about their incentive. In the Simultaneous treatment, the information about the incentive and the quality signal appeared on the same screen. We counterbalanced whether the information about the incentive appeared on the top or the bottom of the

screen, to control for the potential effect of position on the screen on attention. Then, participants were prompted to make a recommendation. We further collected additional measures of beliefs and selfishness using the same measures used in the NoChoice Experiment. At the end of the experiments, we randomly selected 1 out of 10 advisors and sent their recommendation to a client.

D.2 Results

As pre-registered, we merge the data from the NoChoiceSimultaneous experiment with the data collected for the NoChoice experiment, and control for the wave in which the data was collected. The main sample comprises of a total of 276 attentive participants from the NoChoiceSimultaneous experiment and 298 attentive participants from the original wave of the NoChoice experiment, for a total of 574 attentive participants. However, in this experiment, overall, we had much lower data quality than in the prior wave of the NoChoice experiment as well as all the prior experiments. Among those who completed the NoChoiceSimultaneous experiment, 50.3% ($N = 283$) of participants switched multiple times in the multiple price list measure of selfishness, one of our exclusion criteria in the pre-registration. This fraction is much larger than the fraction of inconsistent participants in any of the other study we ran.² Given these differences in data quality, we analyze the data both including and excluding participants who switch more than once in the measure of selfishness.

As shown in Table D.1, participants in the Simultaneous treatment, who received both the information about the incentive and the quality signal on the same screen, were more likely to recommend the incentivized product in cases of conflict than participants in the Assess Quality First treatment. As shown in the table, these participants behaved similarly to those in the See Incentive First treatment. The results are similar both if we include (Column 3) and exclude (Columns 1-2) inattentive participants.

The 276 attentive participants were matched with $N = 28$ clients for the main task; of these 96% followed the recommendation. They were also matched with $N = 28$ clients for the MPL task; of these 79% followed the recommendation.³

²At the time we ran the experiment, Cloudresearch changed some of the features it used to filter participants (<https://www.cloudresearch.com/resources/blog/cloudresearch-is-retiring-the-block-low-quality-participants-option/>) In particular, CloudResearch removed their “Block Low Quality Participants” which is what we have used in all prior experiments. This change resulted in data quality issues as, at the time we ran the study, we could not filter out inattentive participants/BOTs as well as before.

³We also recruited advisees ($N = 28$ for the main task and $N = 28$ for the MPL task that measured moral costs) for the $N = 283$ inattentive participants who switched multiple times in the MPL.

Table D.1: Advisor Recommendations - No Choice (Simultaneous)

	(1)	(2)	(3)
	Main Sample		Including Inattentive
See Incentive First	0.167*** (0.051)	0.155*** (0.050)	0.151*** (0.043)
No Conflict	0.249*** (0.062)	0.226*** (0.061)	0.137** (0.058)
See Incentive First * No Conflict	-0.156* (0.083)	-0.133 (0.083)	-0.109 (0.077)
Simultaneous	0.172** (0.067)	0.174*** (0.066)	0.120** (0.051)
Simultaneous X No Conflict	-0.267** (0.104)	-0.256** (0.101)	-0.083 (0.083)
Incentive for B	-0.149*** (0.037)	-0.156*** (0.036)	-0.163*** (0.031)
Selfishness		0.083*** (0.018)	
Constant	0.745*** (0.077)	0.745*** (0.076)	0.745*** (0.067)
Observations	574	574	883
R^2	0.069	0.104	0.053

Note: This table displays the estimated coefficients from linear probability models on the advisors' recommendations. See Incentive first is a binary indicator coded as 1 for participants who were randomly assigned to see the incentive first. Simultaneous is a binary indicator coded as 1 for participants who saw all information at the same time. Selfishness was elicited at the end of the experiment, using a multiple price list (MPL) with 5 decisions. The variable is a standardized measure of the number of times the advisor chose to recommend the incentivized product in the MPL task. The regression models in columns (1) and (2) restrict the analyses to participants who did not switch multiple times in the MPL. Column (3) includes all participants. The regression includes individual controls for the advisor's gender and age, and a binary indicator for the wave in which participants took part in the experiment. Robust standard errors (HC3) in parentheses

E The Choice Deterministic Experiment

E.1 Experimental Design

The goal of this experiment is to establish whether the behavior of participants in the Choice experiment is affected by our design choice in the main experiment of assigning individuals to their preferred information order with 75% chance. While this design choice allowed us to separate selection from the effect of actually getting flexibility or commitment, it is possible that the presence of uncertainty may have provided participants with an additional excuse to behave self-servingly, affecting both information preferences and subsequent behavior.

In the Choice Deterministic experiment, we replicate the Choice Free treatment from the Choice Experiment and randomly assign participants to one of two treatments that vary whether assignment to the preferred information order occur with 75% chance as in the original experiment, or is certain. We then add the *ChoiceFree-Deterministic* treatment in which advisors know that they will receive information in their desired order with certainty. Comparing these two treatments allows us to understand whether the presence of uncertainty with respect to how advisors received information, conditional on preferring a given order, affected their recommendation behavior.

Procedures. The experiment was conducted on Amazon Mechanical Turk (AMT) and was pre-registered on aspredicted.org (#82298). Participants received \$1 payment for taking part in the experiment and for making a recommendation. On top of that, they received a \$.15 commission for recommending either Product A or Product B. Advisors were informed that one out of 10 advisors would be matched with a client, another AMT participant, and their advice was delivered to them. In the Choice Free-Probabilistic experiment, the procedures were identical to those in the Choice Free treatment of the Choice Experiment. In particular, 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. In the Choice Free-Deterministic experiment, participants were not told that there was a 75% chance that their preference would be implemented. Instead, upon making their choice, advisors proceeded to receive information in their desired order. Upon making their recommendations, we collected additional measures of beliefs and morality using the same measures used in the Choice Experiment. At the end of the experiments, we randomly selected 1 out of 10 advisors and sent their recommendation to a client.

E.2 Results

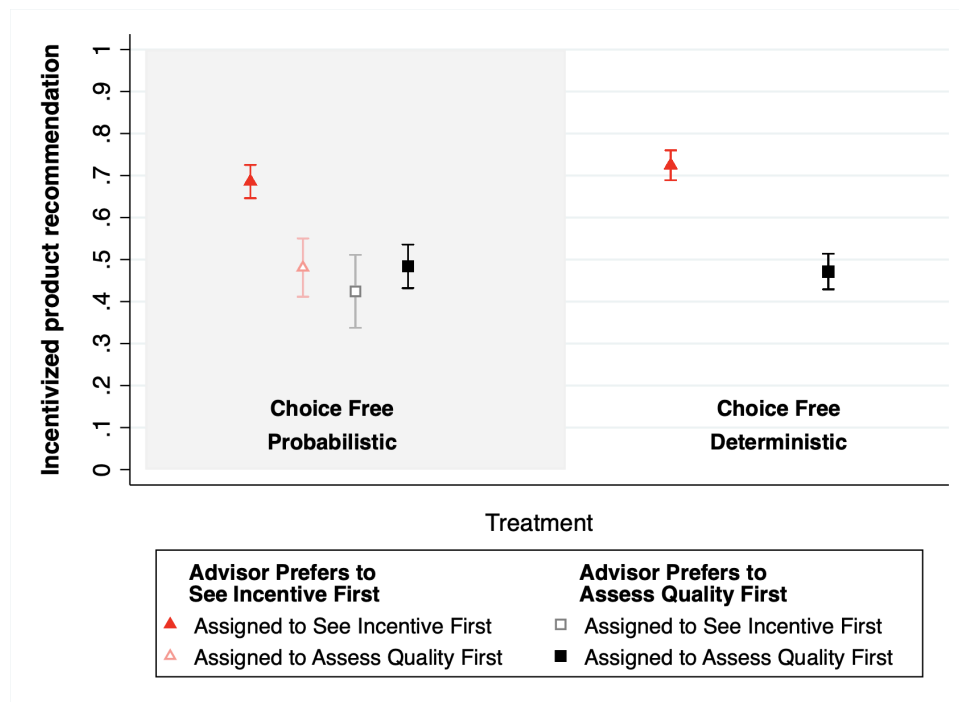
A total of 952 attentive participants completed the experiment; of these 198 participants (20.8%) gave inconsistent responses to the multiple price list measure of morality. Following our pre-registration, we conduct analyses excluding these participants as our main analyses ($N = 369$ participants in the Choice Free-Replication treatment and $N = 385$ in the Choice Free-Deterministic treatment), and also including these participants.

We find that 59.7% of advisors demand to see the incentive first in the Choice Free-Probabilistic treatment and 53.4% in the Choice Free - Deterministic treatment. The decrease in demand is marginally significant (χ^2 -test= 3.09, $p = 0.078$), and consistent with prior work showing that self-serving behavior increases when there is uncertainty (e.g., Haisley and Weber, 2010; Exley, 2015). Including inattentive participants, demand to see the incentive first is 62.9% and 58.7%, respectively, and the difference is not significant (χ^2 -test= 1.66, $p = 0.197$).

Figure E.1. shows advisors' recommendation decisions, when there is a conflict between the signal of quality and the advisor's incentive. The difference in recommendations, depending on advisors' preferences (and when assigned their preference) is similar in the Choice Free-Deterministic and the Choice Free-Probabilistic treatments. A total of 70% of the $N = 223$ participants who preferred and got assigned to see the incentive first recommended the incentivized option in the Choice Free-Replication treatment; this fraction was 76% (out of $N = 281$ participants) in the Choice Free - Deterministic treatment. For those who preferred and got assigned to assess quality first, 55% (out of $N = 132$) and 53% out of $N = 197$) of participants recommended the incentivized option. Further, the figure shows that only 52% (out of $N = 75$) participants who preferred but were not assigned to see the incentive first in the Choice Free -Replication treatment recommended the incentivized option; and 52% (out of $N = 33$) participants who preferred but did not get assigned to see quality first recommended the incentivized option.

Advisors who prefer to see the incentive first (and are assigned their preferred information order) are, on average, 21 percentage points more likely to recommend the incentivized product, as shown in Table E.1. Interactions between the Deterministic treatment and preferences as well as the presence of conflict are not significant. Hence, the results show that recommendation decisions are robust to the probabilistic implementation of advisors' preferences for information order.

At the end of the experiment, we recruited $N = 76$ clients and matched them with 1 out of 10 advisors for the main task; of these 87% followed the recommendation. Advisors were also matched with $N = 76$ additional advisees for the MPL task that measured moral costs; of these 84% followed the recommendation.



Notes: This figure presents the covariate-adjusted recommendations of the incentivized product when there is a conflict between the signal of quality and the advisor's incentive

Figure E.1: Advisors' Recommendations

Table E.1: Recommendations: Assigned Preferences

	(1)	(2)	(3)
	Main Sample		Including Inattentive
Prefer to See Incentive First	0.211*** (0.069)	0.197*** (0.068)	0.200*** (0.064)
No Conflict	0.253*** (0.095)	0.265*** (0.094)	0.253*** (0.091)
Prefer to See Incentive First * No Conflict	-0.099 (0.120)	-0.087 (0.118)	-0.153 (0.113)
Deterministic	-0.015 (0.071)	-0.019 (0.071)	-0.012 (0.067)
Deterministic X No Conflict	0.040 (0.124)	0.030 (0.122)	-0.016 (0.118)
Deterministic X Prefer to See Incentive First	0.102 (0.090)	0.113 (0.088)	0.053 (0.083)
Deterministic X Prefer to See Incentive First x No Conflict	-0.089 (0.154)	-0.103 (0.152)	0.053 (0.145)
Incentive for B	-0.145*** (0.036)	-0.151*** (0.035)	-0.118*** (0.033)
Female	-0.016 (0.036)	-0.002 (0.035)	0.014 (0.032)
Age	-0.002 (0.001)	-0.002 (0.001)	-0.001 (0.001)
Selfishness		0.078*** (0.017)	
Constant	0.616*** (0.082)	0.618*** (0.081)	0.582*** (0.076)
Observations	656	656	832
R^2	0.113	0.141	0.080

Note: This table displays the estimated coefficients from linear probability models on the advisors' recommendations. Deterministic is a binary indicator coded as 1 for participants in the Deterministic treatment. Selfishness was elicited at the end of the experiment, using a multiple price list (MPL) with 5 decisions. The variable is a standardized measure of the number of times the advisor chose to recommend the incentivized product in the MPL task. The regression model in column (3) extends the analyses to included advisors who switched multiple times in the multiple price list eliciting selfishness. The regression includes individual controls for the advisor's gender and age. Robust standard errors (HC3) in parentheses

F Additional Data: Predictions

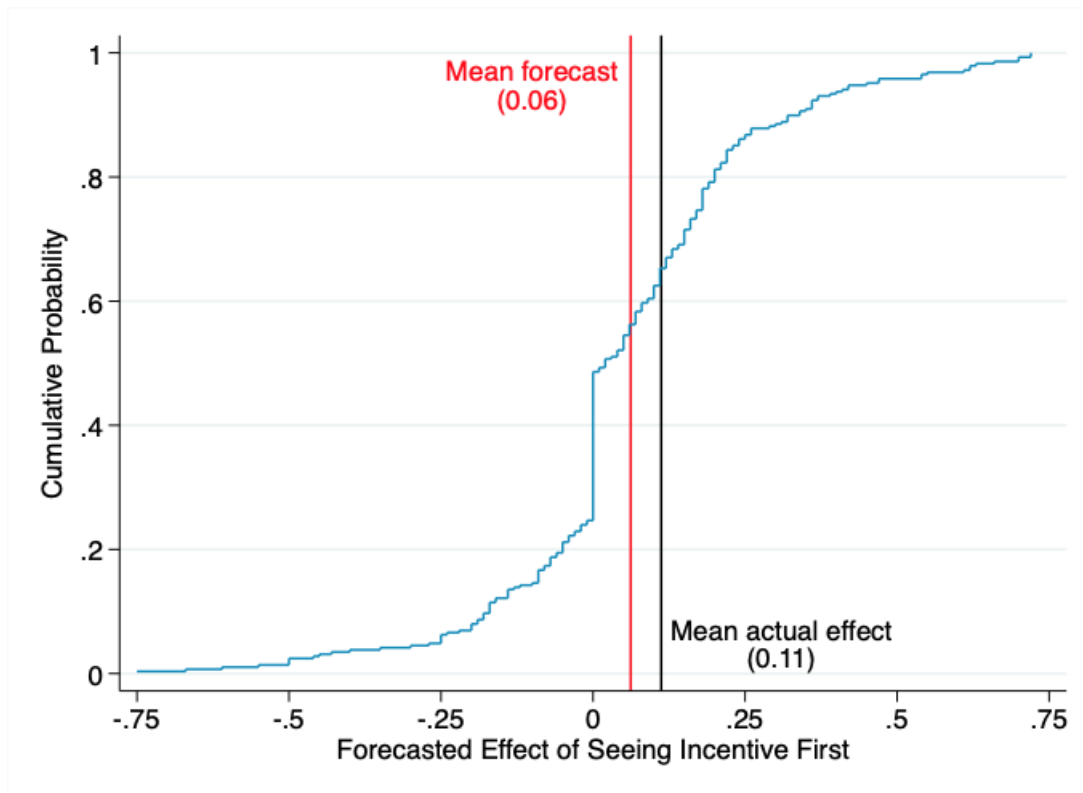
In the Prediction experiment, we recruited forecasters on AMT and asked them to read a summary description of the recommendation decisions advisors made in the Incentive First Costly treatment of the Choice Experiment. A link to the original instruction was provided to participants. We then asked forecasters to predict the recommendation decisions of advisors who choose to see the incentive first. In particular, forecasters were told to consider the recommendation decisions of advisors who chose to see their incentives first. They were asked to estimate the recommendations of advisors who were either assigned to see the incentives first or assigned to assess quality first.

To aid participants in making their predictions, and following the approach of DellaVigna and Pope (2018), participants received 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. For this purpose, we provided forecasters with the fraction of incentivized recommendations in the See Incentive First Costly treatment (AMT-1) of the Choice experiment. 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. Forecasters were paid \$1 and received an additional \$2 bonus if their predictions laid within 5 percentage points of the true value.

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 F.1 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 ($p = 0.395$), which we documented in the See Incentive First Costly (AMT-1) of the Choice exper-

iment. As shown in Figure F.1, 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.



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 F.1: Predicted and Actual Effect of Seeing the Incentive First on Recommendations

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.

G Experimental Instructions

Below we present instructions for the Choice experiment and the IA experiment.

G.1 Choice Experiment

Below we present the screenshots that advisors were presented with in the Choice experiment.

Welcome to the experiment

In today's study, you have been assigned the role of **ADVISOR**.

You will be asked to make a recommendation to another MTurk participant, the **CLIENT**.

At the end of this study, we will randomly choose one advisor out of 10 and give his/her recommendation to a client, who will be then paid accordingly.

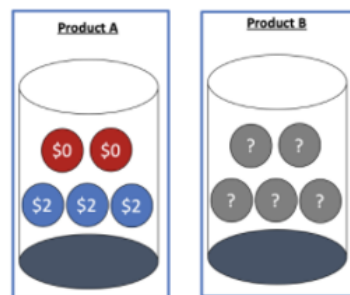
How it works

- You, the **ADVISOR**, will receive information about **two Products, A and B**.
- Your task is to evaluate the products and **recommend one** of the two to the other participant - the **CLIENT**.
- The **CLIENT** will be asked to choose between the two Products (A and B). The clients' choice will affect his/her bonus payment.
- Note that the **client knows nothing about Product A or B**. **The only information he/she will receive about the products is your recommendation.**

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.

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)

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.

Advisor's choice in See Quality First Costly (adjusted accordingly for Choice Free and See Incentive First Costly).

Your payment

- Your task is to recommend either Product A or B to the client.
- You will receive **\$1 for completing this HIT** 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 at random by the computer.

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 receive \$0.05 and 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. If you prefer to learn about the commission before and your preferred option is implemented you will receive an additional \$0.05.

Choice screen

Your Choice

- Recall that you will receive \$1 for completing this HIT 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 receive \$0.05 and 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

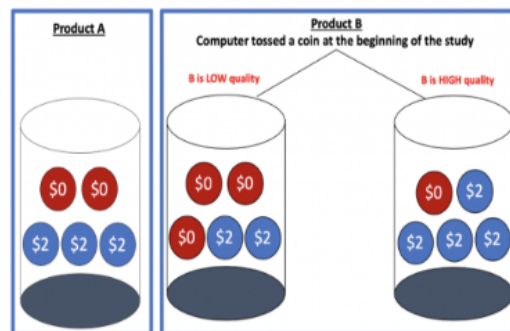
Assignment screen (adjusted accordingly depending on the assignment)

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

Information for advisors who see the incentive after (incentive shown earlier if assignment is to before)

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.
-

Quality signal

• We drew the following payoff (ball) from **Product B**:



This ball will be now placed back into the urn.

Before moving to the next screen, please carefully consider which recommendation you would like to make to the client.

Recommendation decision

Next, we ask you to make a recommendation for your client.

If you recommend Product B, you will receive an additional \$0.15 commission.

Which product do you recommend?

Product A <input type="radio"/>	Product B <input type="radio"/>
---	---

Additional measures of advisors' beliefs and preferences

We will now ask you several additional questions. You can earn an additional payment depending on your responses. Please consider your answers carefully.

After observing the ball, what do you think is the likelihood that the quality of Product B is LOW? (0% means that Product B is extremely *unlikely* to be of LOW quality, whereas 100% means that Product B is extremely *likely* to be of LOW quality)

Please indicate your estimated likelihood by choosing one of the ten options below.

Bonus payment. There are 10 options. If your answer is correct (i.e., if you choose the right range), you will receive an additional **\$0.15 payment.**

To get the bonus, consider your answer carefully.

- The likelihood that **Product B** is of low quality is between **0% and 10%**
- The likelihood that **Product B** is of low quality is between **11% and 20%**
- The likelihood that **Product B** is of low quality is between **21% and 30%**
- The likelihood that **Product B** is of low quality is between **31% and 40%**
- The likelihood that **Product B** is of low quality is between **41% and 50%**
- The likelihood that **Product B** is of low quality is between **51% and 60%**
- The likelihood that **Product B** is of low quality is between **61% and 70%**
- The likelihood that **Product B** is of low quality is between **71% and 80%**
- The likelihood that **Product B** is of low quality is between **81% and 90%**
- The likelihood that **Product B** is of low quality is between **91% and 100%**

Now that you have chosen a bin, what do you believe is the **exact likelihood**? Please enter a number from 0 to 100.

Next, you will complete 2 additional tasks that will ask you to make two additional recommendations. Each of the 2 tasks will have specific instructions and payments. Please read them carefully.

Moral costs

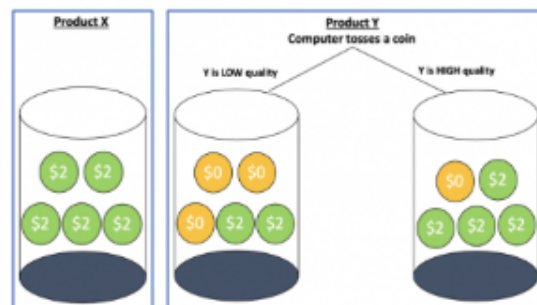
Task #1

- Next, we will show you **5 new sets of Products**.
- For each set, we will ask you to recommend one of two products, X or Y, to an **advisee**.
- The **advisee** is a **different MTurk Worker** than the CLIENT (for whom you made your earlier recommendation).
- At the end of the study, we will randomly select 1 of the 5 sets of products and 1 out of 10 advisors to send the advisee the corresponding recommendation.
- When the advisee receives your recommendation, he/she will be asked to choose between Product A and Product B, without having any information other than your recommendation.

In each of the following 5 sets, **Product X** will change whereas **Product Y** will always stay the same.

In decisions 1 through 5, if you recommend Product Y you will receive an additional \$0.15 payment.

1. If you recommend Product Y for the randomly selected set that is sent to the advisee, you will receive \$0.15.



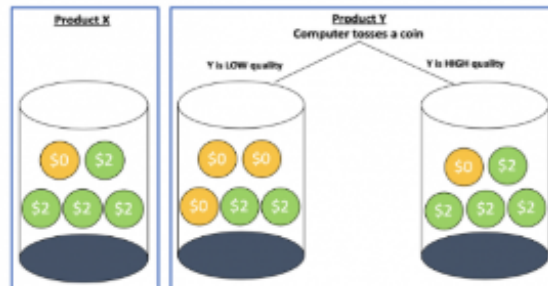
We draw the following payoff (ball)  from Product Y.

Which product do you recommend to the advisee?

Product X

Product Y

2. If you recommend product Y for the randomly selected set that is sent to the advisee, you will receive \$0.15.



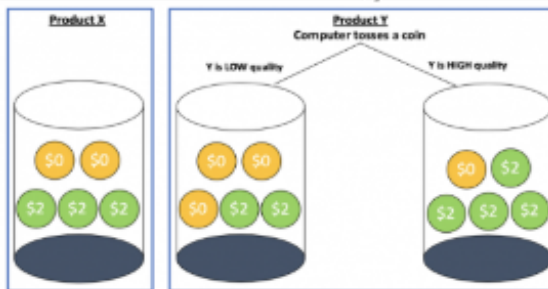
We drew the following payoff (ball)  from Product Y.

Which product do you recommend to the advisee?

Product X

Product Y

3. If you recommend product Y for the randomly selected set that is sent to the advisee, you will receive \$0.15.



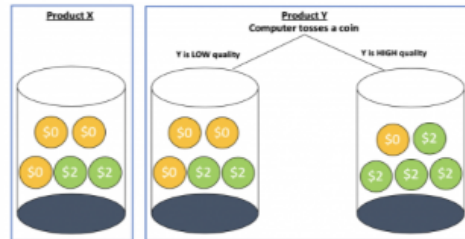
We drew the following payoff (ball)  from Product Y.

Which product do you recommend to the advisee?

Product X

Product Y

4. If you recommend product Y for the randomly selected set that is sent to the advisee, you will receive \$0.15.



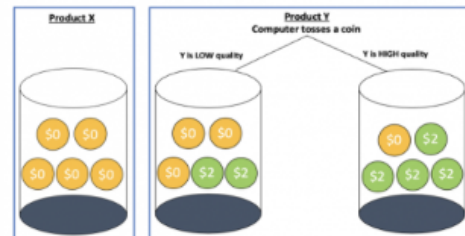
We drew the following payoff (ball)  from Product Y.

Which product do you recommend to the advisee?

Product X

Product Y

5. If you recommend product Y for the randomly selected set that is sent to the advisee, you will receive \$0.15.



We drew the following payoff (ball)  from Product Y.

Which product do you recommend to the advisee?

Product X

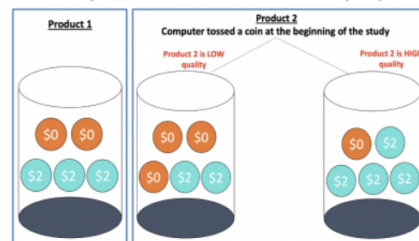
Product Y

Blinding

Task #2.

- For this task, you again will be asked to make a recommendation between two products to another participant, **an ADVISEE**.
- This ADVISEE (another MTurk worker) is different from the one in TASK 1. **At the end of the study, we will randomly select 1 out of 10 advisors to send the advisee the corresponding recommendation.**

The two products, **1 and 2**, are displayed below.



- Product 2 can have **LOW or HIGH quality**, with 50% chance but you don't know if the quality is LOW or HIGH.

To help with your recommendation, you will receive information about the quality of product 2.

- That is, you will be shown a ball (\$0 or \$2), randomly drawn from Product 2, **which will help you infer its quality**. The ball will then be placed in the urn and you will be asked to make your recommendation.

I understand that I will soon receive information about the quality of Product 2

You could receive a **\$0.15 commission**, depending on your recommendation.

- The commission can be for **product 1 OR product 2**, determined at random, and you will learn it before the end of the HIT.

You have the option to learn whether the \$0.15 commission is for product 1 or product 2 **after** you have made your recommendation or **before** making your recommendation.

- **If you choose to see the commission after**, in the next screen you will learn about the quality of Product 2 and will then be asked to make your recommendation. In the following screen, you will learn the commission.
- **If you choose to see the commission before**, in the next screen you will learn about the quality of Product 2 and whether the commission is for product 1 or 2. You will then be asked to make your recommendation.


Which option do you prefer?

Learn whether the commission is for product 1 or 2 AFTER making my recommendation, i.e., in the next screen I will receive information about the quality of Product 2 and then make my recommendation. I will learn the commission in the following screen.

Learn whether the commission is for product 1 or 2 BEFORE making my recommendation, i.e., in the next screen I will receive information about the quality of Product 2, about whether the commission is for Product 1 or 2, and then make my recommendation.

Recommendation decision screen if blinded

We drew the following payoff (ball) from Product 2:



This ball will be now placed back into the urn.

Please carefully consider which recommendation you would like to make to the client.

Which product do you recommend?

Product 1	Product 2
<input type="radio"/>	<input type="radio"/>

Incentive information screen (shown after recommendation for advisors who chose to blind)

The commission is for recommending Product 1.

G.2 Information Architect experiment

Welcome to the experiment

In today's study, you have been assigned the role of **DECISION-MAKER**.

There are two other participants in the role of **ADVISOR** and **CLIENT**.

The **ADVISOR** will be asked to make a recommendation to the **CLIENT**. At the end of this study, we will randomly choose 1 advisor out of 10 to give his/her recommendation to a client, who will be then paid accordingly.

Your decision today may affect the ADVISOR's recommendation and the CLIENT's payment. Your task today is to protect the CLIENT's interest.

How it works

- The **ADVISOR** will receive information about **two Products, A and B**.
- The **ADVISOR's** task is to evaluate the products and **recommend one** of the two to the other participant - the **CLIENT**.
- The **CLIENT** will be asked to choose between the two Products (A and B). The clients' choice will affect his/her bonus payment.
- Note that the **CLIENT knows nothing about Product A or B**. **The only information he/she will receive about the products is the ADVISOR's recommendation.**

- **You, the DECISION-MAKER, can decide how an ADVISOR will receive information that helps them evaluate the two Products, A and B.**

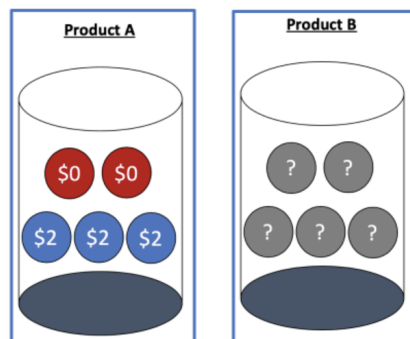
Understanding Question: What is your role and that of others in this study?

- I am the DECISION-MAKER, and in my role I provide advice to another participant in the role of CLIENT.
- I am the DECISION-MAKER, and in my role I assign a product to the ADVISOR (Product A or B)
- I am the DECISION-MAKER, and in my role I will make a decision that may affect the ADVISOR's recommendation and the CLIENT's payment

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. The advisor does not know whether Product B is high or low quality for sure, **but will receive information that will help him/her infer the quality of Product B.**



After receiving the advisor's 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.

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)

What You Know

- The advisor will soon receive more information that will help them 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 client will then will be paid accordingly.

Information for IAs in the IA-client treatment (adjusted accordingly for IA-advisor).

Advisor's payment

- The ADVISOR's task is to recommend either Product A or B to the client.
- In addition to their fixed fee for completing this HIT, the advisor may receive an additional payment depending on their decisions.
- **Depending on which product the advisor recommends, they may receive an additional \$0.15 commission as a bonus payment.**
- The **\$0.15 commission** can be for recommending Product A or B. This has been determined at random by the computer.

Client's payment

- The CLIENT will receive a product recommendation from the ADVISOR.
- **After receiving the advisor's 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.

Your payment

- You will be paid \$1 for completing the study.
- You will receive **an additional \$0.15 payment if the advisor recommends the best product for the CLIENT.**
- That is, if the advisor recommends **the product with more \$2 balls**, you receive a **\$0.15 payment.**

Your choice

As Decision-Maker, you can decide how the advisor will receive information.

- You can choose for the advisor to learn about his/her commission (i.e., whether product A or B yields a \$0.15 commission) **before** or **after** obtaining information that will help the advisor 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 the advisor to learn which product recommendation gives them (the advisor) a **\$0.15 commission (Product A or B) before** the advisor obtains information that helps them infer the quality of Product B

OR

 - I want the advisor to learn which product recommendation gives them (the advisor) a **\$0.15 commission (Product A or B) after** the advisor obtains information that helps them infer the quality of Product B
- Your preferred option will be implemented with 75% chance.

Question: What is your payment today?

- I am paid a \$1 fixed fee.
- I am paid a \$1 fixed fee. On top of that, I receive an additional \$0.15 if the advisor recommends the product that yields them (the advisor) the commission.
- I am paid a \$1 fixed fee. On top of that, I receive an additional \$0.15 if the advisor recommends the best product to the client (i.e., the product with more \$2 balls)

Your Choice

- Recall that you will receive \$1 for completing this HIT and making this choice.
- You will receive an **additional \$0.15 payment if the advisor recommends the best product to the client** (i.e., the product with more \$2 balls).

What do you prefer?

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

Next, you will complete an additional task that will ask you to make some recommendations. The task will have specific instructions and payments. Please read them carefully.

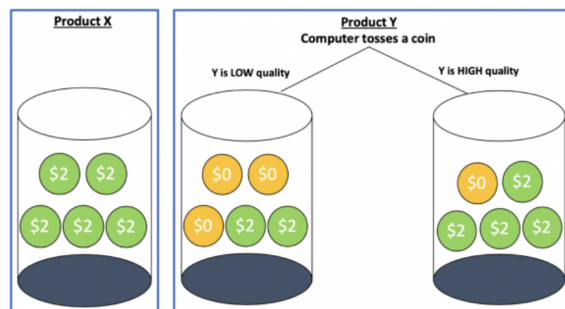
Extra Task

- Next, we will show you **5 new sets of Products**.
- For each set, we will ask you to recommend one of two products, X or Y, to an **advisee**. That is, you will be an advisor.
- The **advisee** is a **different MTurk Worker** than the ADVISOR and CLIENT (for whom you made a decision earlier).
- At the end of the study, we will randomly select 1 of the 5 sets of products and 1 out of 10 participants to send the advisee the corresponding recommendation.
- When the advisee receives your recommendation, he/she will be asked to choose between Product A and Product B, without having any information other than your recommendation.

In each of the following 5 sets, **Product X** will change whereas **Product Y** will always stay the same.

In decisions 1 through 5, if you recommend Product Y you will receive an additional \$0.15 payment.

1. If you recommend Product Y for the randomly selected set that is sent to the advisee, you will receive \$0.15.



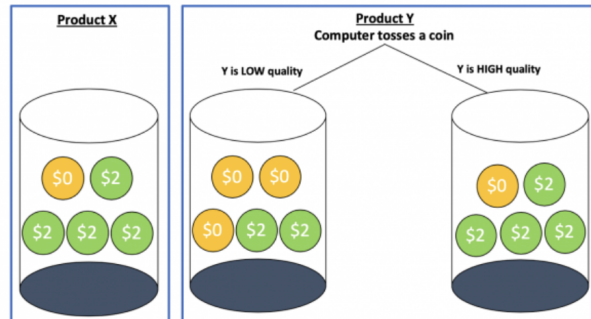
We drew the following payoff (ball)  from Product Y.


Which product do you recommend to the advisee?

Product X

Product Y

2. If you recommend product Y for the randomly selected set that is sent to the advisee, you will receive \$0.15.



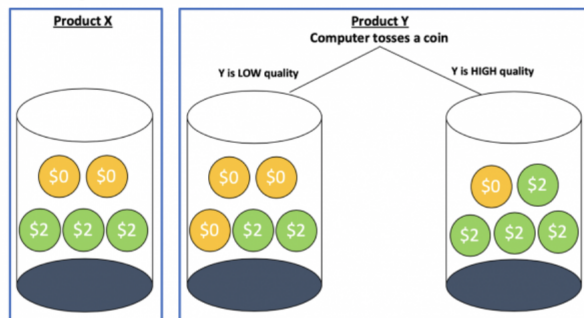
We drew the following payoff (ball)  from Product Y.


Which product do you recommend to the advisee?

Product X

Product Y

3. If you recommend product Y for the randomly selected set that is sent to the advisee, you will receive \$0.15.



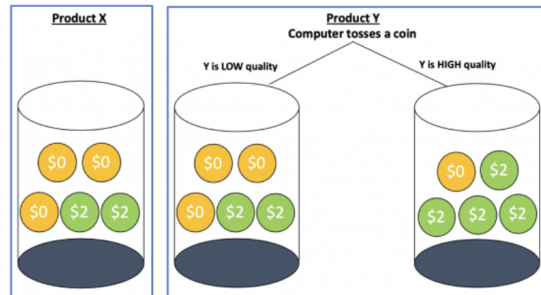
We drew the following payoff (ball)  from Product Y.

Which product do you recommend to the advisee?

Product X

Product Y

4. If you recommend product Y for the randomly selected set that is sent to the advisee, you will receive \$0.15.



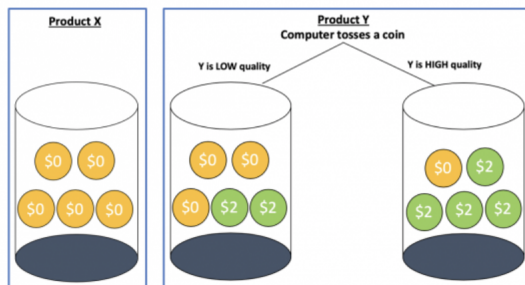
We drew the following payoff (ball)  from Product Y.

Which product do you recommend to the advisee?

Product X

Product Y

5. If you recommend product Y for the randomly selected set that is sent to the advisee, you will receive \$0.15.



We drew the following payoff (ball)  from Product Y.

Which product do you recommend to the advisee?

Product X

Product Y