Persistent Bias in Advice-Giving

Zhuoqiong (Charlie) Chen HIT Shenzhen

Tobias Gesche ETH Zurich

January 2019

Abstract

We show that a one-off incentive to bias advice can have persistent effects. In an experiment, advisers were paid a bonus to recommend to a less-informed client a risky option, relative to its alternatives, which is only attractive to risk-seeking individuals. Afterwards, the advisers learned that they had to choose for themselves and make a second recommendation from the same set of options, this time without the bonus. We find that the bonus biases not only the initial recommendations but also subsequent actions. The advisers who were offered the bonus only for their first recommendation chose the risky option and recommended it a second time up to six times more often than did the advisers in a control group who were never offered a bonus. In an additional treatment, the advisers were informed about this sequence of actions and the one-off nature of the bonus before they made any decisions. Enabling advisers to anticipate the entire sequence of actions does not decrease the bias in the initial recommendation. However, it is effective in eliminating the bonus' persistent effect on the advisers' subsequent own choices and second recommendations. We also present a theory, based on the advisers' (self-) image concerns over appearing biased, which can explain our results.

Keywords: advice-giving, conflict of interest, self-signaling

JEL Classification: C91, D03, D83, G11

Contact: chenzq926@gmail.com and tobias.gesche.job@gmail.com

We thank Björn Bartling, Antonio Cabrales, Lucas Coffman, Erik Eyster, Johannes Kunz, David Laibson, Ferdinand Langnickel, Ryan O. Murphy, Kristóf Madarász, Charles Nossair, David de Meza, Larry Samuelson, Rani Spiegler, Roel van Veldhuizen, Joel van der Weele and several seminar and conference audiences for comments and feedback. Tobias Gesche also acknowledges support by the Swiss National Science Foundation (SNSF Grant #P2ZHP1_171900).

1 Introduction

Giving advice is at the heart of many professions where experts use their knowledge to guide less-informed clients on difficult and risky decisions. However, advisers often face a conflict of interest. Incentives such as sale commissions and kickbacks lead them to ignore clients' actual needs and advertise specific products, most prominently for financial advice (Mullainathan et al., 2012; Malmendier and Shanthikumar, 2014; Egan et al., 2018). Such bias in advice-giving can even be generated more subtly through unconditional gifts (Malmendier and Schmidt, 2017)); however, their consequences are vast. For retirement investment in the US alone, which is only a share of the overall market for advised funds, conflicted advice is estimated to cause a 12% loss over returns for 30-year savings. This corresponds to losses of $17bn. per year (CEA, 2015). In other domains, for example when doctors advise patients on risky treatments, conflicted advice is also a problem (Dana and Loewenstein, 2003; Cain and Detsky, 2008), and the stakes might even be higher, albeit more difficult to quantify. Considering these economic and ethical problems and the fact that disclosure often does not help, removing the cause of the conflict of interest seems to be appealing.[1] In fact, policies that aim at removing and banning adverse incentives for advisers have been suggested or are being considered in various jurisdictions.[2] This paper asks whether such policies can be effective in de-biasing advice and what the obstacles are to achieve this goal.

To answer these questions, we conducted an experiment in which subjects advised others on risky decisions. The subjects who acted as advisers had better information about the risk characteristics of the possible choices on which they advised other subjects who acted as their clients. Some advisers were paid a bonus if they recommended a particularly risky option, whereas advisers in a control condition were not offered this bonus. We first find that this incentive works. Approximately half of the advisers who were offered the bonus recommended the risky option, whereas only a small minority recommended the risky option in the control condition. To determine whether this incentive persists longer, the advisers had to make two additional choices. They first learned that they would also have to choose for

[1] There is currently numerous evidence that disclosing, as opposed to removing, conflicts of interest of experts not only may be ineffective but also may backfire, for example, in Cain et al. (2005), Koch and Schmidt (2010) or Cain et al. (2011). Loewenstein et al. (2014) reviews the psychological literature and mechanisms that underlie these effects; complementing economic accounts are presented in Li and Madarasz (2008), Inderst and Ottaviani (2012), and Gesche (2016).

[2] In the US, laws which would impose a fiduciary duty on retirement advisers are currently being discussed. Such a duty would prevent them from taking side payments which can affect their advice. In the UK, the "Retail-Distribution-Review" which bans commission-based financial advice came into force in 2013. Other European countries differ in the degree to which they regulate incentives which can lead to biased advice. Proposed policies range from banning them altogether (e.g., Netherlands), for some services (e.g., Italy) or not at all and just requiring disclosure (e.g., Germany). The MiFID II-directive by the European Union, which became effective in 2018, also calls for avoidance of conflicts of interest in financial advice-giving (see Article 34(3)(c-d) of the directive).

themselves from the same set of options but now without any bonus attached. After this step, they also learned that they had to issue a second recommendation to another client who had not received advice before, again, with the bonus removed. We find that the advisers who were previously offered the bonus were three times more likely to choose the risky option for themselves than the advisers in the control condition where no bonus was ever paid. The removed bonus also affected second recommendations. The advisers who had previously been exposed to the bonus were six times more likely to recommend the risky option to another client than the advisers in the control condition.

These findings are consistent with a simple theory that we present. Its underlying reasoning is based on two main notions. The first notion is that being influenced by a bonus, i.e., not recommending what one considers appropriate advice, is deemed to be immoral. Consistent with this, almost half of the advisers in our experiment who were initially offered the bonus did not recommend the risky option. The second notion is that people want to avoid the inference that their initial advice was biased. To signal one's own moral integrity, advice has to be unaffected by the bonus. This therefore requires consistency in advice-giving, even when this entails repeating biased advice after the bonus was removed. Consistent with such a mechanism, we estimate that approximately 40% of the advisers whose initial advice was biased by the bonus recommended the risky option again.

Our experiment also allows us to ask how advisers determine what appropriate advice is and the implications of this when there is a conflict of interest. If advisers linked their own choices to what they consider appropriate advice, they also have to choose accordingly to avoid signaling that they gave biased advice. In line with this, we find that for the advisers whose initial advice was biased by the bonus, a sizable fraction – again approximately 40% – also chose the risky option for themselves. These results speak against an alternative reasoning where advisers could have self-servingly assumed that their clients are risk-seeking.

To further investigate the mechanisms that underlie the persistent biases that we document, we also conducted an additional treatment. Similar to the previous treatment with a bonus, the advisers in this treatment were also paid the same bonus to recommend the same risky option and then, with the bonus removed, had to choose for themselves and recommend again. However, instead of learning about these stages one after another, they were told about this sequence of decisions and the temporary nature of the bonus from the beginning. We find that this possibility to anticipate the consequences of biased advice did not weaken the bonus' initial effect. When advisers could anticipate the upcoming decisions, their first recommendations were as biased as when a bonus was paid and they could not anticipate this. However, we find that anticipation is effective in preventing a persistent bias: When they could be anticipated, advisers' own choices and second recommendations were not different to the control condition in which no bonus was ever paid. These findings therefore demonstrate the potential of ex-ante considerations to prevent long-lasting, adverse effects of conflicts of interests in advice-giving.

2 Related literature

Recent observations on real-world adviser behavior resonate strongly with our findings. Foerster et al. (2017) use observational data on approximately 6,000 Canadian financial advisers and more than 580,000 of their clients. They show that not the clients' personal characteristics but simple fixed effects for the individual advisers explain most of the variation in how risky the clients' investment portfolios are. By using the same data set, Linnainmaa et al. (2018) report that recommendations to clients resemble the choices that these advisers make for themselves. These advisers chose the same return-chasing and actively managed funds that their respective clients held. The advisers chose the same funds although these investments were riskier than other investments and performed worse than the market average. In addition, the advisers held these underperforming portfolios even after they left the industry. Our work explains how sales commissions (which are typically paid for by selling such funds) can cause such patterns. However, we do not use observational data from the field. Rather, we employ an experimental approach that allows us to exogenously vary the exposure to the bonus. Thus, the findings in our experiment cannot be explained by an alternative theory in which advisers self-select into compensation schemes that cater to their personal preferences or beliefs.

Closely related to our findings is also an experiment by Gneezy et al. (2016). For a single recommendation, they report a relatively low bias in their "after" condition. In it, the advisers saw all available options and then had to consider which option to recommend. Afterwards, they learned that they could earn a bonus if they subsequently recommended a specific option. In their "before" treatment, this order is reversed: The advisers learned about the bonus before they could see the available options, could then consider which option to recommend, and then had to make an actual recommendation. They find that the bias towards recommending the option with the bonus is greater when advisers learn about the bonus before getting further information about the options than when they learn about it afterwards. This finding is consistent with the explanation that we offer. This explanation captures the notion that changing advice signals the fact that one has given biased advice. In their setup, this occurs when advisers first see the options, consider what to recommend, and then change the actual advice when they learn about the bonus for recommending a specific option (i.e., in their "after" treatment). If, in contrast, advisers first learn about the bonus, they can direct their initial consideration for what they want to recommend towards the option with the bonus. Then, their initial consideration and their actual recommendation for this option are aligned so that recommending it does not signal a bias (i.e., giving biased advice is facilitated).[3] Building on Gneezy et al.'s experimental framework and extending it to a dynamic setting, we i) formalize this explanation and present further evidence which is consistent with it, ii) show that it can lead to persistent bias in advice-giving when the conflict of interest is removed, iii) show that the underlying behavioral mechanism can also affect advisers' own choices, and iv) demonstrate that allowing advisers to foresee the consequences of their behavior can limit such persistent effects.

These results also connect to previous findings showing that role-induced dispositions lead people to align their judgment. In a classic study, Festinger and Carlsmith (1959) demonstrate that paying subjects to report favorably about an unpleasant task improves their subsequent evaluation of this task, relative to when they were not paid. A similar spill-over occurs in experiment by Loewenstein et al. (1993) who had subjects acting in the fictitious role of plaintiff or defendant in a legal case. This role affected what the subjects considered to be a fair settlement value towards the interest of the respective role that they temporarily took. In the same experimental setting, Babcock et al. (1995) report that subjects find it more difficult to agree on a settlement value when they knew their role before learning about the case's details as opposed to when they first learn about the case and then whether they are plaintiff or defendant. These differences in what is considered to be a fair settlement value can be explained by a desire to minimize cognitive dissonance (Festinger, 1957).[4] In our case, this arises if people want to perceive themselves as unbiased advisers but their own actions indicate the opposite.

To avoid experiencing cognitive dissonance, own actions and opinions can be adjusted, even if this is costly and effectively leads to self-deception (Trivers, 2011). In fact, Schwardmann and van der Weele (2016) show that when tasked to convince other people of their own ability, subjects overstate

[3] In the context of a joke-writing-contest where the referees could be bribed, Gneezy et al. (2018) report a similar effect. In their "KeepWinner"-treatment, contestants submit their jokes together with bribes whereas in their "KeepWinnerDelayed"-treatment, contestants submit bribes after their jokes were submitted and initially screened. They find a higher distortion in referee judgments in the former treatment, similar to the above-described "before"/"after"-comparison. In another experiment, Bicchieri et al. (2019) elicit subjects' beliefs about the lying behavior of others, either before or after subjects knew that they themselves would have the possibility to lie. If elicited before, subjects state a lower belief about others' lying rates and also engage in less lying than when beliefs are elicited after they know that they could lie themselves.

[4] Similarly, Konow (2000) finds that subjects' perceived contributions and entitlements to a collectively generated surplus shift systematically, depending on the roles they had in a subsequent dictator game in which this surplus was split. For economic models of cognitive dissonance, see Akerlof and Dickens (1982) and Rabin (1994).

their own ability in subsequent private self-assessments, despite the fact that such overstating is costly to themselves. A similar finding comes from Mijović-Prelec and Prelec (2010). They report a similar pattern where subjects self-deceive to avoid the inference that they made an error in a classification task and how such behavior can be derived through a self-signaling mechanism (see Bodner and Prelec, 2003; Bénabou and Tirole, 2004, 2011).[5] Our work takes up on these insights and investigates their implications in the context of advice-giving.

We therefore also contribute to the literature on the moral underpinnings of sender behavior in strategic communication (e.g. Gneezy, 2005; Sutter, 2009; Rode, 2010; Inderst et al., 2018). Instead of a sender-receiver game, we examine the lasting consequences of conflicts of interests in the related but different situation of advice-giving (which is particularly important in the context of credence goods, see Kerschbamer and Sutter, 2017).[6] Thus, this work also links to the recent literature on the adverse effects of bonus payments (Christoffersen et al., 2013; Bénabou and Tirole, 2016) and how underlying resulting conflicts of interests shape the self-perception and attitudes of the people exposed to them, for example, in the financial industry (Burks and Krupka, 2012; Cohn et al., 2014; Zingales, 2015). However, our findings come from a neutral framing and relate to biased advice-giving more generally.

Finally, we also connect to the general literature on moral reasoning and economic behavior. A central principle therein is the notion that people care about being perceived as moral persons and that their own actions signal their underlying motivations (Bodner and Prelec, 2003; Bénabou and Tirole, 2004), in particular, their own moral values (Bénabou and Tirole, 2011; Benabou et al., 2018). Although such image concerns can refer to both one's social or self-image, self-image alone can steer moral behavior. This applies to, for example, non-maximal lying to uphold the illusion of being honest (Mazar et al., 2008), being less likely to follow an economic incentive to administer electric shocks to others while seeing oneself on a video screen (Falk, 2017), or customers avoiding purchases under pay-what-you-want schemes to avoid appearing to be greedy to themselves (Gneezy et al., 2012). As a key theoretical result, we show how image concerns can cause the effects of conflicts of interest to persist. Empirically, our experiment shows support for this in a setting where social-image concerns are

[5]See Kunda (1992) for a discussion on how cognitive dissonance and self-deception relate. Falk and Zimmermann (2017a,b) show another instance of costly consistency. They report that subjects forfeit opportunities to improve their accuracy in estimation tasks in order to signal ability to a principal or themselves.

[6]In sender-receiver games, the sender can observe an external event and then communicate it to the receivers via a message which, given a defined language, can be "true" or "false". In advice-giving, the sender's message is not about such an objectively observable state. Rather, it is a suggestion on what ought to be done by an expert who has better information than the receiving party.

minimized. Thus, we present evidence which indicates that self-image concerns can also matter in the domain of advice-giving.

A related branch of this literature, which was recently summarized by Gino et al. (2017), shows that information is often not processed in an objective manner if this threatens a person's self-image. Instead, people act as "motivated Bayesians" who instrumentalize uncertainty, ambiguity, and the tendency to err in a self-serving way (e.g., Dana et al., 2007; Haisley and Weber, 2010; Exley, 2016; Exley and Kessler, 2018). Importantly, this includes the formation of beliefs about other people and their preferences to accommodate one's own selfish actions (Di Tella and Pérez-Truglia, 2015). Our theory and experiment also allows us to investigate the relevance of such reasoning in the context of biased advice. We then find that our results cannot be explained solely by self-serving beliefs about others' risk preferences. Rather, our results indicate that advisers often base their recommendation on their own preference (see also Mullen et al., 1985; Faro and Rottenstreich, 2006; Ifcher and Zarghamee, 2018).

3 Experimental design and procedures

To empirically investigate whether and how conflicts of interest persist, we conducted a controlled experiment that builds on the setup by Gneezy et al. (2016). At the beginning of the experiment, the subjects were directed to cubicles with computer screens on which instructions were provided. They were informed that they would earn a show-up fee of GBP 5.00 and that there would be further possibilities to earn money. The subjects participated in a session in which either everybody was an adviser or everybody was a client.

In the adviser sessions, the subjects were informed that they would act as advisers for clients who would participate in a future client session at the same laboratory. The advisers then learned that they had to recommend one of three risky choices, which were referred to as Option A, B, and C, to their clients. First, they received the following information about the three options: *"Each option will earn different monetary payoffs. Option A presents a possibility to earn a high or low payoff, depending on luck. Option B adds the possibility to earn some amount between the high and low payoff, Option C increases that possibility."* They also learned that clients would receive only this piece of information but that they, as advisers, would receive additional information.

The advisers received the additional information via screens that informed them about the payoffs associated with the options and how the payoffs would be determined. This information was also given to them on a sheet of paper with further examples and explanations. The advisers could keep this sheet

Table 1. Description of the investment options as shown to advisers (but not to clients)

Die equal to:	**Option A**	**Option B**	**Option C**
1 or 2	lottery: GBP 20 or 0	safe payment: GBP 12	safe payment: GBP 12
3 or 4	lottery: GBP 20 or 0	lottery: GBP 20 or 0	safe payment: GBP 8
5 or 6	lottery: GBP 20 or 0	lottery: GBP 20 or 0	lottery: GBP 20 or 0

Note: In the above, "lottery" is a fair coin toss in which "Heads" wins GBP 20 and "Tails" nothing.

as a reference throughout the study. Table 1, which was also on this sheet, summarizes how the payoffs for each possible option are determined. A person who has chosen an option would roll a fair, six-sided die. Depending on the chosen option, the person then either receives a safe payment (e.g., GBP 12 if Option C is chosen and the die shows a one) or has to play a binary lottery by tossing a coin that earns GBP 20 with "Heads" and nothing with "Tails" (e.g., if Option B is chosen and the die shows a five). As a consequence, the payments for all options are realized independently.

Note that a choice among the three options allows a categorization of the underlying risk preferences. If one compares Options A and B, only a person who is willing to give up a safe payment of GBP 12 to gamble with an expected payment of GBP 10, i.e., a risk-seeking individual, chooses Option A. Conversely, Option C is preferred to Option B only by a person who wants to sacrifice an expected payment of GBP 10 for a safe payment of GBP 8. Thus, only a risk-averse individual should choose Option C. Accordingly, choosing Option B requires an individual to be neither sufficiently risk-averse nor sufficiently risk-seeking, i.e., such a choice reflects approximate risk-neutrality. Considering this ordering, we will henceforth refer to Options $A/B/C$ as the risky/neutral/safe options, respectively.

Structure of the experiment. The adviser sessions proceeded along the following five steps:

Step 1 – First recommendation R1: After the advisers had studied the instructions, they were asked to make a recommendation to clients. For this, they had to write on a piece of paper that they recommend their client to choose either Option A, B or C. They were then instructed to put this paper in an envelope that was collected by an experimenter and put into a box.

Step 2 – Own choice O: When all advisers had written down their recommendation R1 and all envelopes were collected, they were informed that they would now have to choose one of the three options for themselves. The advisers were previously not informed about this step. The procedure was the same as for issuing advice to clients. The subjects had to write their choice on a letter and put it in an envelope. An experimenter came by and collected the envelopes and put them in a separate box.

Step 3 – Second recommendation R2: After they had made their own choice in O, the advisers were asked to make a second recommendation to a different client from the client in R1 who would not receive further advice. Again, this step was not announced before the preceding step had finished, and recommendations were made by writing them on letters that were then collected.

Step 4 – Questionnaire: When all recommendations from R2 were collected, the advisers had to complete a short on-screen questionnaire that elicited personal information. It also included a short question on the advisers' general willingness to take risks.

Step 5 – Sampling & Payoff: At the end of each session, one envelope was sampled from each of the boxes for R1, O, and R2 to become effective. For each of the sampled envelopes from the R1 and R2 boxes, the corresponding recommendation sheet was passed to a different client in a future client session. For the sampled envelope from the O box, the corresponding adviser actually earned the payoff from his/her choice. Thus, for every adviser session, there was one adviser who actually got paid out the choice from O and two clients in the later client session who received advice from the advisers' recommendations in R1 and R2 of that adviser session.

For the sampled envelopes, the corresponding cubicle number (which was written on the envelopes) but not the recommendation or choice itself was read aloud so that a subject knew whether his/her envelope was sampled. This procedure of sampling and calling the cubicle numbers at the end of the experiment was explained before the advisers made their respective recommendations and choices. Payment was then conducted by calling the subjects, one by one, to the laboratory's exit where they were paid in private according to their choices (if a subject's envelope was chosen for O, the subject also had to toss the die and coin to determine the corresponding payoff).

Treatments NO BONUS and BONUS: The above structure describes the experimental procedure in our baseline treatment to which we will refer to as NO BONUS. A second treatment, called BONUS, features an incentive of GBP 3 to recommend the risky Option A in R1. The advisers in BONUS learned about this incentive after they were informed that they had to give advice but before they saw the sheet with the detailed information about the investment options. This bonus was only paid in the subject's first recommendation R1. On the screens that explained the O and R2 tasks, it was clearly stated that there would not be any additional bonus for choosing or re-recommending Option A. This one-off bonus in R1 is the only difference between BONUS and NO BONUS.[7] Advisers in BONUS

[7]The reason for paying the bonus if Option A was recommended and not only if clients actually chose it, is that this keeps the setup simple and efficient with regards to biased advice, the focus of our study. For the same reason, the bonus

Table 2. Treatment overview

Advisers:	**NO BONUS**	**BONUS**	**ANTICIPATE**
Get bonus	no	3 GBP for A in R1	3 GBP for A in R1
Learn about O	after R1	after R1	before R1
Learn about R2	after O	after O	before R1
No. of observations	51	48	50

Note: when advisers learned about O and R2, they also learned that there was no bonus in these stages.

learned about this incentive after having been informed that they had to give advice but before seeing the sheet with the detailed information about the investment options. This bonus was only paid for subject's first recommendation R1. On the screens which explained the O and R2 tasks, it was clearly stated that there would not be any additional bonus for choosing or re-recommending Option A. This one-off bonus in R1 is the only difference between BONUS and NO BONUS.[8]

Treatment ANTICIPATE: To investigate the effect of the advisers' knowledge of upcoming actions and to rule out alternative explanations, we conducted a third additional treatment. It largely resembled BONUS, i.e., it promised a bonus of GBP 3 if an adviser recommended Option A in R1. However, the advisers in ANTICIPATE not only were told about the bonus in R1 before they made their first recommendation but also received additional information. That is, they learned about O and R2 and that the bonus would be removed for these two steps. They got this additional information on a separate screen that appeared after they learned about the general setup of the advice-giving situation and the bonus but before they were actually asked to make their first recommendation. Thus, the difference in ANTICIPATE relative to BONUS is the information about the upcoming decisions in O and R2 and the one-off nature of the bonus. Table 2 summarizes the treatments and their differences.

Client sessions: In the week after the adviser sessions, additional subjects from the same subject pool participated in further sessions. In these sessions, the subjects acted as clients, and each client received one of the sampled recommendations for R1 or R2 from each of the 11 previous adviser sessions. Thus,

of 3 GBP was relatively high. As shown in the results section, even this bonus did not succeed in leading almost half of the adviser to recommend Option A while the relevance of biased advice is undisputed. Also note that even if the payment of the bonus were conditional on a client's choice for Option A, this would require to recommend this option.

[8]Since advisers' payoffs in BONUS do not depend on the clients' decisions, they were not explicitly informed about whether clients would learn about the bonus. Also, none of the advisers asked for this information even though they were encouraged to ask clarifying questions. Clients were informed about the bonus when they received a recommendation R1 from an adviser who had been in the BONUS treatment.

there were a total of 22 clients. After reading the recommendations, the clients made their choices and were then paid accordingly. The advisers knew about this structure. In this paper, we focus on them.[9]

Verifiability: To ensure that the advisers believed that if their recommendation was randomly chosen it would become effective and shown to a client, we used the following procedure in all treatments. We allowed the advisers to voluntarily sign their recommendations and address the envelopes to themselves. It was explained to the advisers that if their recommendation was chosen to be shown to a client, the sheet would be signed by the respective client. In case that the corresponding adviser had provided us with his or her address on the envelope, the adviser would then receive a copy of the signed recommendation by mail. Note that the client only received the recommendation letter, not the envelope; therefore, the adviser was anonymous to the client (the advisers knew this). The subjects were informed about this option before they made their first recommendation and were reminded of it before the second recommendation. It was also emphasized that this option was entirely voluntary. Together with the announcement of sampled recommendations' cubicle numbers (see Step 5 above), the advisers knew whether to expect a letter. Accordingly, the advisers knew that the experimenters were pre-committed to actually show the sampled advice letters to clients.[10]

General procedures: Throughout the experiment, we enforced a strict no communication policy. We conducted 11 adviser sessions, each with 11 to 17 (in total 149) subjects who acted as advisers. The advisers earned on average GBP 6.89 (around USD 9.50 when the experiments occurred), and no session lasted longer than 45 minutes. All subjects were students across several degrees and fields of studies (see Table C.1 in Appendix C for the descriptive statistics). All the instructions, with the exception of the paper reference sheet that explained the options, were shown on a computer interface (which was programmed with zTree, see Fischbacher, 2007).[11] For screenshots and copies of the instructions, see Appendix D. The experimental sessions were conducted in January 2016 (treatments NO BONUS and BONUS) and April 2018 (treatment ANTICIPATE) at the London School of Economics' Behavioural Research Lab with subjects from its pool. Before the experiments, the principal design and research

[9] Given that we have six relevant conditions (recommendations from R1 vs. R2 and BONUS vs. NO BONUS vs. ANTICIPATE) and that, due to our random sampling procedure, the 22 clients are not balanced across these conditions, there is not much analysis which can be done due to limited statistical power.

[10] We also checked whether the regression results for R1 and R2 as presented in Section 5 are affected by the inclusion of controls for whether a recommendation letter was signed and/or an envelope was addressed. They are unaffected and none of these additional control variables is significant.

[11] Using a computerized interface allowed to track subjects' progress to provide them with the necessary information in time, for example about O and R2 in NO BONUS and BONUS *after* the respective preceding stages had been completed. The use of handwritten letters allowed to implement the verifiability mechanism described above.

questions of the study were submitted to the school's research ethics committee as a part of the (successful) procedure to obtain its approval.

4 Behavioral mechanism and predictions

In this section, we describe how through a simple model, a preference to appear as an unbiased adviser can lead to a persistent bias in advice and in an adviser's own choices. The model is based on an adviser ("he") who advises a client ("she") and is concerned with what his current actions reveal about his past motivations to give advice. Specifically, it assumes that an adviser's overall utility consists of the following three elements: 1) a standard vNM-utility derived from (expected) monetary payoffs; 2) the psychological or material costs of not giving appropriate advice; and 3) the diagnostic dis-utility of learning from one's own current actions that previous advice was biased.

Although the first element is standard, the second reflects advisers' uneasiness to recommend something that they do not consider to be appropriate advice. For example, an adviser might think that a recommendation for a particular choice is suited to his client because, given the adviser's belief about the client's preferences, this would be the client's preferred choice if she had the same information as the adviser. Not recommending this preferred choice then creates costs because the adviser has not acted in the client's best interest.[12] One way to determine what constitutes appropriate advice is by acting based on a belief about the client's preferences. Importantly, this includes the possibility to form a motivated belief about the client's preference that accommodates biased advice. Another way to determine appropriate advice is via an adviser's own preferences, i.e., by the answer to the question "What would I choose if I were in the client's position?"[13] Our theory and experiment allow us to explore the different implications that these possible lines of reasoning have. However, for the principal mechanism of repeated biased advice that we explore, it only matters that the costs of recommending something inappropriate exist, irrespective of whether (in)appropriate advice is determined through an adviser's own preferences or an independent, possibly motivated, belief concerning the client's preferences.

[12] In fact, for many adviser-client-relations such as doctors and patients, lawyers and clients, and several situations of financial advice-giving, there is a fiduciary duty which legally requires the adviser to act in the client's best interest.

[13] That people use this question to decide for others, in a variety of domains has recently been shown by Ifcher and Zarghamee (2018) who call it the "Golden Rule". More generally, it is a robust psychological fact that people base their inferences about others' preferences on their own (Marks and Miller, 1987), in particular for risk preferences (Faro and Rottenstreich, 2006). Even though initially coined by Ross et al. (1977) as a "false consensus effect", the falsity of estimating others' preferences based on one's own is not evident. Works by Hoch (1987) and Dawes (1990) demonstrate that often, such projection is not just statistically correct; they also show that people can often improve their accuracy in predicting others' preferences by relying more strongly on their own. Engelmann and Strobel (2000) show that subjects do so when they are incentivized to make accurate predictions.

The third element, diagnostic dis-utility, arises from image threats that advisers experience when their actions reveal that they have given biased advice, i.e., through a self-signaling mechanism. In contrast to the costs of giving inappropriate advice, this diagnostic dis-utility only occurs to an adviser after he has biased his initial recommendation, at the point when his later actions – his own choice or second recommendation – indicate exactly this fact to him. The important implication of such an inference is that advisers can only uphold a positive image of themselves as long as they do not take actions that are incompatible with this notion. This relates our model to similar approaches that also feature, in addition to an outcome utility, a diagnostic (dis-)utility (Bodner and Prelec, 2003; Bénabou and Tirole, 2004, 2011; Grossman and van der Weele, 2017). In our context, it is derived from the image concerns that an adviser experiences upon learning that prior advice was biased.

Together, these three components then have implications for how and, most importantly, for how long conflicts of interest can affect advisers' actions. To see this, consider an adviser who gave biased advice. Thus, his costs of not giving appropriate advice were smaller than the (material) benefit that he obtains from giving some other, thus biased, advice. If the adviser is also sufficiently concerned about his image, he then needs to continue to give the same biased advice again, even when the conflict of interest has disappeared. The reason is that to entertain the notion that the initial advice was unbiased, it should be unaffected by the presence of any external incentive. However, changing advice after the conflict of interest disappears signals the opposite. If an adviser's own preferences determine what constitutes appropriate advice, there can be further consequences. In case he needs to make a choice for himself, the adviser is put on the spot if he has previously given biased advice. This is because not choosing as he initially recommended then also signals his previous bias. If advisers' image concerns are sufficiently high, this might lead them to choose consistently, thereby effectively biasing their own choices. Therefore, a behavioral trait that generally seems to be desirable – a preference to be perceived as unbiased – can lead to persistent biases.

4.1 A model of persistent biases in advice-giving

We now present a simplified model of the above reasoning that is closely connected to our experimental design and that allows us to derive hypotheses from it. A full-fledged, more general version is contained in Appendix A. We denote the first recommendation r_1, the adviser's own choice by o, and the second recommendation by r_2, where $r_1, o, r_2 \in \{A, B, C\}$. Denote by $r^*, o^* \in \{A, B, C\}$ what an adviser considers to be appropriate advice and the choice that he would prefer for himself if there were no other motives, respectively. For advisers in the BONUS and ANTICIPATE-treatments, there is a bonus b associated with recommending Option A in the first recommendation R1. Reflecting the three elements of an adviser's utility function as described above, the resulting payoff when taking an action $a \in \{r_1, o, r_2\}$ is given by

$$U(a \mid h_a, r^*) = \underbrace{u(v(a))}_{\text{vNM}} \quad \underbrace{-\kappa \cdot \mathbb{1}[a \in \{r_1, r_2\} \text{ and } a \neq r^*]}_{\text{costs of giving inappropriate advice}} \quad \underbrace{-\lambda \cdot \Pr[r_1 \neq r^* \mid h_a, a]}_{\text{diagnostic dis-utility}} \tag{1}$$

where $h_a \in \{\emptyset, r_1, (r_1, o)\}$ is the history of the adviser's choices. Specifically, $h_a = \emptyset$ if $a = r_1$; $h_a = r_1$ if $a = o$; and $h_a = (r_1, o)$ if $a = r_2$.

Each term in (1) corresponds to an element in the psychological mechanism that we introduced above. The first term of the payoff function (1) is a vNM utility upon receiving a payment v and we assume $u(0) = 0$. For the first recommendation in our experiment R1, $v(r_1)$ equals $b \geq 0$ (with $b = 3$ GBP in BONUS and ANTICIPATE and $b = 0$ GBP in NO BONUS) if $r_1 = A$ while $v(r_1) = 0$ if $r_1 \in \{B, C\}$. For a choice o made for the adviser himself, $v(o)$ represents the corresponding certainty equivalent. In the second term, $\mathbb{1}[\cdot]$ is an indicator function that takes the value of 1 if the statement in the bracket is true and 0 otherwise; $\kappa \geq 0$ is thus the dis-utility that one experiences by giving inappropriate advice. In the third term, $\lambda \geq 0$ captures the dis-utility the adviser experiences once he learns that the previous advice was biased, i.e., that he recommended something that he does not consider to be appropriate advice ($r_1 \neq r^*$). Such an inference is captured by the posterior probability of such an event, given the adviser's history h_a that precedes his current action a. This probability is therefore calculated from the perspective of an outside observer who only sees an adviser's actions but not his preferences r^* and o^* (which follow known distributions with full support). Also note that, by definition, we have $\Pr[r_1 \neq r^* \mid h_a = \emptyset, a] = 0$ as image concerns are backward-looking.

To see the implication of the above-described setup, first regard advisers who consider Option A to be appropriate to recommend ($r^* = A$). They therefore always recommend it in R1 and R2, regardless of whether there is a bonus or not. In BONUS, there are also advisers who recommended Option A in R2 only because they were biased by the bonus in R1 (i.e., for them $r^* \neq A$ holds). Although they do not consider this option to be appropriate, they can mimic the advisers who actually consider Option A to be appropriate advice by also recommending it in R2. They do this because different recommendations in R2 than in R1 would reveal that they gave a biased recommendation in R1. Thus, with history h_{r_2} containing $r_1 = A$, it holds that

$$\Pr[r_1 \neq r^* \mid h_{r_2}, r_2 = A] < \Pr[r_1 \neq r^* \mid h_{r_2}, r_2 \neq A] = 1. \tag{2}$$

Formally, such behavior can be shown to be part of an equilibrium in a (self-)signaling game. In it, biased advisers can pool with unbiased advisers by re-recommending Option A. Furthermore, one can show that such an equilibrium always exists and also that it is the only one, given the above assumptions (see Appendix A).[14] Therefore, image concerns create an incentive for an adviser to not recommend what he considers to be appropriate advice in R2 after he gave biased advice to earn the bonus in R1. When there is no bonus, this is different as no incentives to bias the advice exist; therefore, image concerns of being perceived as having given biased advice do not matter.

A similar reasoning can be applied regarding advisers' own choices. Suppose advisers determine appropriate advice based on what they prefer for themselves. Unbiased advisers who actually consider Option A to be appropriate advice should then recommend this option in R1 and choose it in O. Biased advisers who do not prefer this option and have recommended it in R1 only for the bonus are then tempted to not choose it for themselves. However, choosing differently in O signals that their first recommendation was biased, while choosing consistently allows them to pool with unbiased advisers. Formally, this corresponds to the expression in (2), except that r_2 is replaced by o. Again, one can show that such behavior is part of the only equilibrium of this situation in which biased advisers pool with unbiased advisers (see Appendix A). Thus, image concerns can create an incentive for advisers to choose Option A in O although they do not prefer it – a situation that does not occur when there was never a bonus that could lead to biased advice.

4.2 Predictions for the treatments BONUS vs. NO BONUS

We now detail the above reasoning and present predictions for the comparisons of our treatment. We start with the first recommendation stage R1. Since this is the first action that an adviser takes, it does not have signaling value regarding past behavior in all three treatments (i.e., for R1 we have

[14] We use Bayesian Nash Equilibrium (BNE) to derive predictions. The full model in the appendix also provides formal definitions and proofs. For ANTICIPATE and the case that advisers could form a multi-stage plan of (potentially) mutually dependent actions in R1, O, and R2 (see subsection 4.3), we use Perfect Bayesian Equilibrium (PBE).

$h_{r_1} = \emptyset$). This implies that image concerns do not matter. In the BONUS treatment, advisers are paid for recommending Option A. Their payoff function (1) is given by

$$U(r_1 \mid h_{r_1}, r^*) = u(v(r_1)) - \kappa \cdot \mathbb{1}[r_1 \neq r^*], \tag{3}$$

with the history $h_{r_1} = \emptyset$ and payoffs $v(A) = b$ and $v(r_1) = 0$ for recommending anything else. Therefore, in addition to the advisers who actually think that Option A is appropriate (those with $r^* = A$), some advisers might be induced to recommend this option to earn the bonus although they do not consider the option to be appropriate to recommend. This happens when the costs of giving inappropriate advice are low relative to the pecuniary utility associated with the bonus, i.e., if $\kappa < u(b)$.

In NO BONUS, however, there is no pecuniary gain to issue any specific recommendation. Absent other motives, only the costs of issuing inappropriate advice remain. Therefore, advisers' payoff function reduces to the second term in (3) so that only advisers who think that Option A is appropriate recommend it. Thus, the following prediction emerges:

Prediction 1. *In R1, advisers recommend Option A more often in BONUS than in NO BONUS.*

We now turn to O, where advisers make a choice for themselves. Therefore, the costs κ of giving inappropriate advice play no role. In the BONUS treatment, there is a cost of being perceived or perceiving oneself as biased since previous advice in R1 could have been biased by the bonus. Accordingly, advisers maximize the following payoff function:

$$U(o \mid h_o, r^*) = u(v(o)) - \lambda \cdot \Pr[r_1 \neq r^* \mid h_o, o] \tag{4}$$

where $h_o = r_1$. If advisers determine appropriate advice according to their own preference (that is, if $r^* = o^*$ holds), then an unbiased adviser's own choice in O and his previous recommendation in R1 should coincide.[15] In the presence of image concerns, this has implications for advisers who initially recommended Option A in R1 just because of the bonus. When they choose differently in O, they signal that their initial advice was biased (see (2) with r_2 replaced by o). Biased advisers can avoid this negative signal if they also choose Option A for themselves; thus, they mimic the advisers who actually prefer it. However, this choice leads to a loss in expected pecuniary utility because they choose Option

[15] We show in the appendix that there is no "reverted" signaling equilibrium in which these advisers choose something different in O, just due to image concerns of being perceived as biased.

A instead of their truly preferred non-A choice. They make this choice only if the image costs λ are high relative to this loss. This is represented by condition

$$\lambda > \underline{\lambda} \equiv \frac{u(v(o^*)) - u(v(o = A))}{1 - \Pr[r_1 \neq r^* \mid h_o, o = A]} \tag{5}$$

where $o = A$, while $o^* \in \{B, C\}$ for such advisers, and history h_o contains $r_1 = A$.

In NO BONUS, there are no image concerns of having issued biased advice before since the advisers could not have been biased by the bonus. Their payoff function reduces to only their pecuniary utility, the first term in (4). Accordingly, all advisers in this treatment just choose their preferred option in this treatment. Thus, if their own preference determines what constitutes appropriate advice, we obtain the following prediction:

Prediction 2a. *In O, if advisers determine what appropriate advice is based on their own preferences ($r^* = o^*$), they choose Option A more often for themselves in BONUS than in NO BONUS.*

If Prediction 2a were wrong, this could have two reasons. The first reason is that what advisers considers to be appropriate advice is independent of their own preferences (i.e., $r^* \perp\!\!\!\perp o^*$ holds). For example, advisers whose advice in R1 was affected by the bonus might have formed a self-serving belief about the client's risk preferences. This would allow them to rationalize their first recommendation for Option A and lower the costs of of giving inappropriate advice. Generally, if appropriate advice is independent of advisers' own preferences, their own choices do not have diagnostic value, i.e., $\Pr[r_1 \neq r^* \mid h_o, o]$ does not vary with o. The same consequence emerges in NO BONUS, here because $\Pr[r_1 \neq r^* \mid h_o, o] = 0$ is a constant and independent of the conditioning variables. Thus, advisers in both treatments choose o to maximize only the first term in (4), that is, they choose their preferred option. This then yields the following, alternative prediction:

Prediction 2b. *In O, if advisers determine what appropriate advice is independently of their own preferences ($r^* \perp\!\!\!\perp o^*$), they choose Option A as often for themselves in BONUS as in NO BONUS.*

The other reason why Prediction 2a might be wrong is simply that advisers do not have sufficiently high image concerns, i.e., $\lambda < \underline{\lambda}$ holds for too many advisers. Absent sufficiently strong image concerns, the initial, biased recommendation should not affect advisers' own choices. To distinguish between these possibilities, one can turn to the second recommendations. If image concerns are absent or too low, the preceding biased advice should not have an effect in R2. However, if image costs do matter, they predict treatment differences in this stage – irrespective of how appropriate advice is determined.

To see this, note that in R2, an adviser's own pecuniary utility is unaffected by what is recommended as there is no bonus in any treatment. However, the previous recommendation might have been biased so that image concerns matter. Thus, not recommending as one has recommended in R1 signals a previous bias, as shown in (2). In addition, there are still costs of giving biased advice. Thus, the adviser's utility for recommendation r_2 and the associated history $h_{r_2} = (r_1, o)$ is given as follows:

$$U(r_2|h_{r_2}, r^*) = -\kappa \cdot \mathbb{1}[r_2 \neq r^*] - \lambda \cdot \Pr[r_1 \neq r^* \mid h_{r_2}, r_2] \tag{6}$$

An unbiased adviser should then just recommend what he actually considers to be appropriate and, therefore, should repeat his initial advice. In particular, the advisers who truly prefer to recommend Option A re-recommend this option.[16] This means that to not be perceived as biased, advisers who have previously been biased by the bonus also have to re-issue the same advice. Thus, when their costs of giving inappropriate advice are small relative to their image costs, they mimic the behavior of advisers who truly consider Option A to be appropriate. Formally, the condition for acting this way is

$$\frac{\lambda}{\kappa} > \nabla \equiv \frac{1}{1 - \Pr[r_1 \neq r^*|h_{r_2}, r_2 = A]} \tag{7}$$

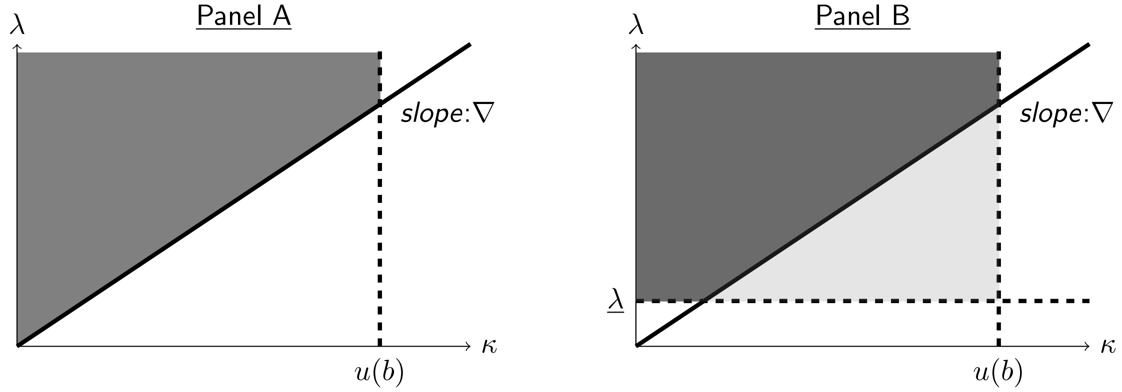
where history h_{r_2} contains $r_1 = A$. As a consequence, advisers whose initial advice was biased and for whom the above applies re-recommend Option A in R2, although they do not consider it to be appropriate advice and even though there is no bonus (anymore) for this recommendation.

In NO BONUS, there was no incentive to bias advice so that image concerns of having given biased advice cannot play any role. Therefore, advisers now maximize only the first term of the payoff function (6). Accordingly, only the costs of giving inappropriate advice matter. Option A is then only recommended by the advisors who already recommended it in R1 because they genuinely consider this option to be appropriate. This yields the following prediction:

Prediction 3. *In R2, advisers recommend Option A more often in BONUS than in NO BONUS.*

[16] As for the own choice O we show in Appendix A that within our model, there is no "reverted" signaling equilibrium in which unbiased advisers' actions are affected by image concerns.

Figure 1. Persistent bias in the (κ, λ)-space



Note: Values of κ and λ which imply persistent bias in R2 if own choices do not have diagnostic value ($r^* \perp\!\!\!\perp o^*$, dark trapezoid in Panel A) and if own choice do have a diagnostic value ($r^* = o^*$, dark pentagon in Panel B). The rectangle formed by the union of the dark grey pentagon and the light grey triangle in panel B depict values which imply persistent bias in O.

Figure 1 visualizes the above reasoning. It depicts the parameter constellations that underlie persistent bias in recommendations and advisers' own choices in BONUS. First, to give biased advice in R1, the costs of doing so κ have to be less than the utility from the bonus $u(b)$, which is depicted by the vertical dashed lines in both panels. For biased advice to be issued again in R2, λ/κ must also be over the threshold ∇ as defined in (7), which is depicted in the figure by the diagonal line. The gray trapezoid in Panel A then depicts the parameter constellation for repeated biased advice when an adviser's own choices are independent of what constitutes appropriate advice.

If an adviser's own choices constitute what is considered to be appropriate advice, Panel B applies. This is because the choices in O then have diagnostic value regarding an adviser's inferred bias. In this case, advisers who have given biased advice in R1 also choose the same for themselves in O if their image concern exceeds $\underline{\lambda}$, as defined in (5). Thus, the rectangle drawn by the two dotted lines in Panel B denotes the parameter constellations for persistent bias in O. For persistent bias in R2, only the advisers who chose consistently in O have not yet revealed themselves to be biased and can therefore continue to behave as if they truly consider Option A to be appropriate. Thus, in addition to the condition displayed by the diagonal in Panel A, the parameter value must also lie in the reactangle in Panel B. Consequently, advisers with parameter combinations in the small white triangle above the diagonal in Panel B reveal themselves to be biased by choosing inconsistently between O and R1. The advisers who exhibit persistent bias in R2 (and O) then have parameter combinations that lie in the dark pentagon.

Our experimental design and the above analytical framework so far enable us to yield two main insights. First, by finding support for Predictions 1 and 3, we can detect a persistent bias in advice-giving caused by image concerns. Second, by testing Prediction 2a, we can investigate this effect in detail. If Prediction 2a is correct, this suggests that appropriate advice is determined based on advisers' own preferences. When advisers' own choices and appropriate advice are not related, e.g., through self-serving beliefs about the client's preferences, we expect Prediction 2b to be confirmed instead.

4.3 Predictions for the treatments BONUS vs. ANTICIPATE

To examine what moderates persistent bias and to rule out alternative explanations, we conducted the ANTICIPATE treatment. The only difference to BONUS is that it gave advisers the possibility to anticipate the upcoming decisions that they had to make before they issued their first recommendation. Depending on whether the advisers in our experiment consider their three decisions to be dynamic choices or not, there are two possibilities. The first possibility is that they ex-ante consider the decisions for R1, O, and R2 to be a *one-stage* static decision with three elements (r_1, o, r_2). Since image concerns are assumed to be backward-looking image concerns, this means that advisers do not factor in image concerns when they make such a one-stage decision – similar to advisers who bias their first recommendation in BONUS. The second possibility is that advisers consider the decisions for R1, O, and R2 to be one sequence of *multi-stage* decisions with three interacting elements. Thus, when making the decision for R1, they can anticipate backward-looking image concerns, potentially triggered by their action in O and R2. Formally, in both cases, advisers choose (r_1, o, r_2) to maximize the following compound utility:

$$\begin{aligned}\sum_{a\in\{r_1,o,r_2\}} U(a \mid h_a, r^*) \quad = \quad & \underbrace{u(v(r_1)) - \kappa \cdot \mathbb{1}[r_1 \neq r^*]}_{\text{R1}} \\ & \underbrace{+u(v(o)) - \lambda \cdot \Pr[r_1 \neq r^* \mid h_o, o]}_{\text{O}} \\ & \underbrace{-\kappa \cdot \mathbb{1}[r_2 \neq r^*] - \lambda \cdot \Pr[r_1 \neq r^* \mid h_{r_2}, r_2]}_{\text{R2}}. \end{aligned} \tag{8}$$

If advisers in ANTICIPATE evaluate their action ex-ante and consider each of them to be a separate one-stage decision, this separation means that for each action, no prior history is evaluated. Formally, in their utility function (8), $h_{r_1} = h_o = h_{r_2} = \emptyset$ then holds and there is no dis-utility from learning that they have given biased advice in R1 because $\Pr[r_1 \neq r^* \mid h_a = \emptyset, a] = 0$ holds for all $a \in \{r_1, o, r_2\}$. In consequence, advisers decide what to recommend for R1 without worrying about its possible implications for what to choose in O and what to recommend in R2. Such behavior would be consistent with previous findings which show that when subjects are given the possibility to ex-ante consider a sequence of decisions (e.g., by using the strategy method) they act less pro-socially than when making decisions sequentially (for an overview, see Cooper and Kagel, 2016).

In this case, the decision process for R1 in ANTICIPATE resembles the decision process in BONUS (i.e., a simple trade-off between the costs of giving inappropriate advice and earning the bonus). Thus, there is no difference in actions for R1 between these two treatments. However, if the possibility to anticipate upcoming actions leads advisers to ex-ante consider them as separate decisions, this makes different predictions for R2 in these two treatments. This is because when backward-looking concerns do not matter in ANTICIPATE, then there is no pressure to act consistently in R2. Thus, no persistent bias occurs in this treatment (in contrast to BONUS, see above). By analogous reasoning for O in ANTICIPATE, the same applies to advisers' own choices if they have diagnostic value (i.e., if Prediction 2a is true). To the extent that advisers follow through with their plans, the following prediction emerges:

Prediction 4a. *If choices in ANTICIPATE are ex-ante considered to be one-stage decisions, then*

- *in R1, advisers recommend Option A as often in ANTICIPATE than in BONUS,*
- *in O, if Pred. 2a is true, advisers choose Option A less often in ANTICIPATE than in BONUS,*
- *in R2, advisers recommend Option A less often in ANTICIPATE than in BONUS.*

The predictions differ in the second case, when advisers can evaluate their actions ex-ante and consider them to be part of a multi-stage decision in which future backward-looking image concerns are anticipated. Such anticipation then creates an additional cost of initially giving biased advice in R1. This is because advisers then factor in that giving such advice "forces" them to either bear the costs of choosing and re-recommending sub-optimally in the upcoming decisions or to suffer image costs from acting inconsistently and thereby revealing their bias in R1. Formally, an adviser then chooses (r_1, o, r_2) to maximize (8) where $h_o = r_1$ and $h_{r_2} = (r_1, o)$. The utility function (8) then captures not only the benefits and costs of the bonus in R1 but also, simultaneously, the additional costs via anticipated image concerns in O and R2. These costs can be anticipated and factored in for ANTICIPATE, but not for BONUS. As the bonus is constant across these two treatments, the following prediction emerges:[17]

[17] Predictions for R2 are not as clear-cut as for R1 because of two counter-veiling effects: On the one hand, the effect on R1 means that fewer advisers recommend Option A in this stage. This decreases the share of those who might have to later make persistently recommendations in R2 to prevent signaling their bias (a decrease on the extensive margin). On the other hand, those who give biased advice in R1 even though they factor in the costs of later recommending in R2 are exactly those for whom these costs are low. They are therefore more likely to act persistently biased once they have given an initial biased advice (an increase on the intensive margin). If own choices determine appropriate advice ($r^* = o^*$), the same reasoning prevents clear-cut predictions for O without making further assumptions.

Prediction 4b. *If choices in ANTICIPATE are ex-ante considered to be a multi-stage decision advisers recommend Option A more often in BONUS than in ANTICIPATE for R1.*

The above predictions for ANTICIPATE are based on advisers who either consider their decision to be a one-stage or multi-stage process, who ex-ante form a plan (for R1, O, and R2), and who then follow through with it. Of course, there is also the possibility that advisers do not follow through with this plan. In this case, they do not anticipate the image costs ex-ante, but when they make own choices or recommend a second time, these costs kick in and affect their decision. In the most extreme case, advisers completely abandon their initial plan. This disregard of initial plans means that the decision situations are then effectively a series of one-shot situations, with a similar prediction as in BONUS. However, if the possibility to anticipate has any effect on advisers' ex-ante considerations and subsequent behavior, then Prediction 4a or 4b follow, depending on whether image costs are factored in ex-ante or not. In addition, if either of these two predictions is found to be true, this also allows us to rule out some alternative mechanisms.

4.4 Comparison to alternative mechanisms

One alternative mechanism that can cause some of the above-described effects is based on anchoring (Tversky and Kahneman, 1974). In it, the advisers who responded to the bonus and recommended Option A in R1 may "anchor" on this option. That is, they consider this option to be the reference option and may then stick to it, even if the bonus is removed. However, such an explanation predicts the same behavioral patterns between BONUS and ANTICIPATE. This is because both of these treatments feature the same bonus for the first recommendation, and before they make any subsequent choice, advisers learn that the bonus disappeared. In consequence, advisers anchor on Option A should do so in both of these treatments. This is different under image concerns (see Prediction 4a and 4b).

A related, somewhat more conscious, mechanism is a "cue"-effect. In it, the bonus is perceived as a signal about an option's quality or what the experimenter wants the advisers to recommend (Zizzo, 2010; de Quidt et al., 2018). If such a bonus-induced cue affects advisors' perceptions of the options, it can then continue to influence decisions in O and R2, even after the bonus disappears. In our model, this would correspond to a shift of r^* and o^* towards Option A, due to the bonus attached to this option. As with anchoring however, the underlying bonus structure is the same in both BONUS and ANTICIPATE so that no differences between these treatments are predicted.

Table 3. Predicted treatment differences for different behavioral mechanisms

	BONUS minus NO BONUS		BONUS minus ANTICIPATE		
	Image concerns	Anchoring & cue-effects	Image concerns		Anchoring & cue-effects
			One-stage	Multi-stage	
R1	$+$	$+$	0	$+$	0
O if $r^* \perp\!\!\!\perp o^*$	0	$+$	0	0	0
O if $r^* = o^*$	$+$	$+$	$+$	n.a.	0
R2	$+$	$+$	$+$	n.a.	0

Note: Predicted differences in the share of advisers who recommend or choose Option A in BONUS relative to the other treatments. $+$ and 0 present a positive or no differences in treatment comparison; n.a. denotes cases where no clear-cut prediction can be made without further assumptions (see footnote 17).

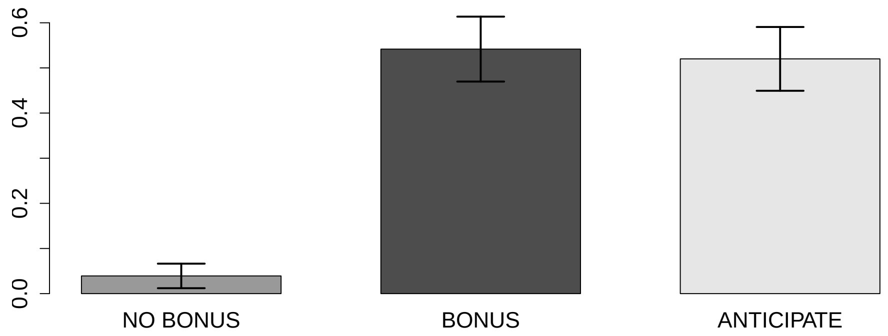
Table 3 summarizes how different behavioral mechanisms make different predictions regarding how often Option A is recommended or chosen and how this differs across the treatment in our experiment. Each prediction follows the reasoning of the mechanism presented in this section and can also be derived from the full model (see Table A.1 in Appendix A). The predicted treatment differences are shown for each action by receivers in our experiment and for O depending on whether advisers' own preferences o^* determine appropriate advice r^* or whether these two elements are independent. The following section reports the results from our experiment which can be used to discriminate between the theories.

5 Results

5.1 Results for the first recommendations (R1)

We start with presenting our results and the tests of our predictions for the first recommendation. This is where our main treatment manipulation occurred. In the treatments BONUS and ANTICIPATE, advisers were paid a bonus to recommend Option A. Accordingly, we expect some advisers to react to this incentive and recommend the risky option more frequently in these treatments than in NO BONUS where no such bonus was paid. Figure 2 portrays the differences in the recommendations for Option A across treatments (for the distributions of adviser actions over all options, see Figure C.1 in Appendix C). In fact, only 3.9% of the advisers in NO BONUS recommended Option A in their first recommendation, whereas more than half of all advisers, 54.2% in the BONUS-treatment and 52.0% in the ANTICIPATE-treatment, recommended this option. These increases relative to NO BONUS are

Figure 2. Advisers' first recommendations (R1) for Option A over treatments (bars depict standard errors)



highly significant (Fisher exact test, BONUS vs. NO BONUS: $p < 0.001$, ANTICIPATE vs. NO BONUS: $p < 0.001$; all tests reported here are two-sided tests). In contrast, the share of recommendations for Option A in the two treatments that paid a bonus do not differ significantly (Fisher exact test, BONUS vs. ANTICIPATE: $p = 0.843$).

We also estimate the following regression model, which includes additional control variables:

$$\mathbb{1}[r_{1,i} = A] = \alpha + \beta \cdot BONUS_i + \gamma \cdot ANTICIPATE_i + \boldsymbol{\delta} \cdot \mathbf{c}_i + \epsilon_i \tag{9}$$

In the above, the dependent variable is an indicator that takes a value of one if subject i's first recommendation $r_{1,i}$ was for Option A. $BONUS_i$ and $ANTICIPATE_i$ are dummies that indicate whether this subject was assigned to the respective treatments as opposed to NO BONUS, the baseline. The vector $\mathbf{c}_i$ collects the control variables that indicate the subject's age, gender, monthly available budget, region of origin, the highest degree that the subject holds or pursues and the field of study.

Table 4 presents the results of this linear probability model, first without controlling for the subjects' characteristics and then with such controls. Again, it is shown that the bonus leads to an increase of approximately 50 percentage points in the probability of recommending Option A in the treatments BONUS and ANTICIPATE. This increase also reflects the previous nonparametric results and is not significantly different between these two treatments, as documented by the corresponding F-tests. The effect of the BONUS-treatment is consistent with Prediction 1, while the equally strong effect of the bonus in ANTICIPATE supports Prediction 4a. The fact that the bonus affects the recommendation equally strongly in both treatments, although its one-off nature and upcoming choices were known in ANTICIPATE, speaks against alternative Prediction 4b.

Table 4. Effect of bonus on advisers' first recommendations in R1

	(1)	(2)
Dependent variable	$r_1 = A$ (first recommendation for Option A)	
BONUS	0.502***	0.448***
	(0.078)	(0.086)
ANTICIPATE	0.481***	0.499***
	(0.076)	(0.086)
Constant (NO BONUS)	0.039	-0.086
	(0.027)	(0.228)
F-test: BONUS=ANTICIPATE	0.050	0.240
Controls	no	yes
Estimation method	OLS	OLS
Data used from treatments	B,A,N	B,A,N
Observations	149	149

Note: Regression results with robust standard errors in parentheses; significance levels against the (two-sided) null of a zero effect: * p<0.10, ** p<0.05, *** p<0.01. The dependent variable is always a dummy indicating a recommendation for Option A in R1. The main independent variables are dummies indicating whether the adviser was in treatment B(=BONUS) or A(=ANTICIPATE) and whether Option A was recommended in the first recommendation R1; N(=NO BONUS) is the reference category. Additional independent variables control for advisers' age, gender, monthly available budget, region of origin, study degree and field of studies.

Note that our results in this stage also feature another important insight: Almost half of the advisers in the BONUS and ANTICIPATE treatments (45.8% and 48.0%, respectively) did *not* recommend Option A, although they were offered money to do so. This feature is consistent with the notion that there exist nonpecuniary costs of giving such advice and that for a considerable fraction of advisers, these costs outweighed the pecuniary utility of the bonus.

5.2 Results for advisers' own choices (O)

For the advisers' own choice, no bonus was paid to the advisers in any condition. Figure 3 displays their choices for Option A across treatments. In the baseline NO BONUS, we observe that 9.8% chose Option A for themselves. This share is comparable to the results of Holt and Laury (2002) who find that 6% to 8% of their subjects exhibit risk-seeking preferences. In BONUS, the advisers were *previously* offered the bonus for their first recommendation. The share of the advisers who chose Option A for themselves is 27.1%, which is almost three times as many advisers as in NO BONUS. This 17.3 percentage-point increase is statistically significant (Fisher exact test: $p = 0.036$). The previously offered bonus therefore continues to affect choices in this treatment where subsequent actions could not be anticipated. This is different in the ANTICIPATE treatment. In it, just 8.0% recommended Option A. This is significantly less than in BONUS (Fisher exact test: $p = 0.016$) but not significantly different from the rate in NO BONUS (Fisher exact test: $p = 1.000$).

Again, these findings are also mirrored in a regression analysis. For this, we replace the dependent variable in regression model (9) with a dummy that indicates whether an adviser chooses Option A for himself. Columns 1 and 2 in Table 5 report the corresponding results without and with added control variables. The results are very similar and show that having been offered a bonus for recommending Option A persistently affects the choices of the advisers who could not anticipate this stage but not for the advisers who could anticipate it. We therefore regard Prediction 2a as supported by our results for BONUS, while the results for ANTICIPATE are consistent with Prediction 4a.

Given these findings, it is helpful to recall the mechanism that underlies our reasoning concerning a persistent bias. It argues that if advisers base what they consider to be impartial advice on their own preferences, then they have to act according to their biased, previous advice to not signal the fact that they were biased. Therefore, the root cause of the persistent effect on the adviser's own choice is that the bonus led advisers to recommend Option A in the first recommendation. This initial bias then affects the advisers' subsequent own choice O in BONUS.

To investigate the mediating effect of the first recommendation, we also estimate the above regression model when an indicator for whether the first recommendation was for Option A (i.e., the dependent variable from model (9)) is included as an additional independent variable. If the bonus' lasting effect on advisers' own choices worked via the initial recommendation, its effect should be captured by the coefficient on this additional regressor. The results in the third column of Table 5 shows that exactly this occurs. The previously positive and statistically significant coefficient for BONUS essentially becomes zero and insignificant, while the coefficient on $r_{1,i} = A$ takes up all its explanatory power. However,

Figure 3. Advisers' own choices (O) for Option A over treatments (bars depict standard errors)

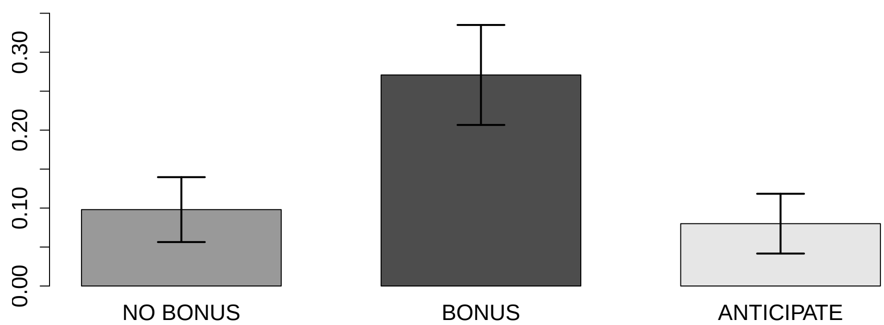


Table 5. Effect of bonus on advisers' own choices in O

	(1)	(2)	(3)	(4)	(5)
Dependent variable			$o = A$ (own choice for Option A)		
BONUS	0.173**	0.181**	0.024		
	(0.077)	(0.084)	(0.076)		
ANTICIPATE	-0.018	-0.009	-0.184***		
	(0.057)	(0.066)	(0.067)		
$r_1 = A$			0.351***		
			(0.078)		
$\widehat{r_1 = A}$ (via BONUS)				0.384**	
				(0.150)	
$\widehat{r_1 = A}$ (via ANTICIPATE)					-0.119
					(0.145)
Constant (NO BONUS)	0.098**	-0.018	0.012	-0.390	0.177
	(0.042)	(0.175)	(0.149)	(0.316)	(0.168)
F-test: BONUS=ANTICIPATE	6.390**	5.600**	7.360***	-	-
Controls	no	yes	yes	yes	yes
Estimation method	OLS	OLS	OLS	2SLS	2SLS
Data used from treatments	B,A,N	B,A,N	B,A,N	B,N	A,N
Observations	149	149	149	99	101

Note: Regression results with robust standard errors in parentheses; significance levels against the (two-sided) null of a zero effect: * p<0.10, ** p<0.05, *** p<0.01. The dependent variable is always a dummy indicating a choice for Option A in O. The main independent variables are dummies indicating whether the adviser was in treatment B(=BONUS) or A(=ANTICIPATE) and whether Option A was recommended in the first recommendation R1; N(=NO BONUS) is the reference category. Additional independent variables control for advisers' age, gender, monthly available budget, region of origin, study degree and field of studies (columns 2–5). Estimates are based on OLS (columns 1–3) or 2SLS where a recommendation for Option A in R1 is instrumented by assignment to treatment B/A and data from treatment A/B is not used (column 4/5, respectively).

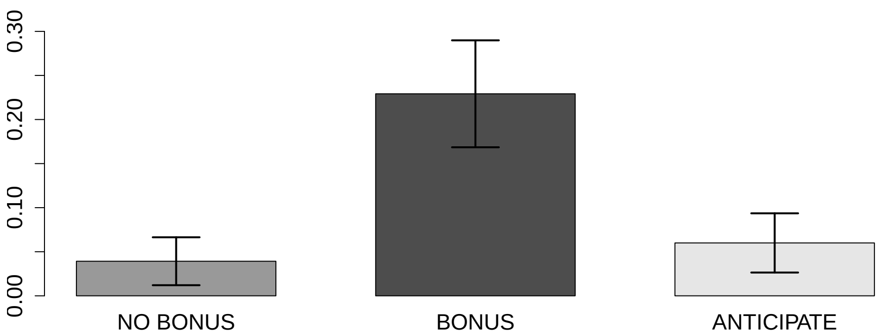
even if one controls for the initial recommendation for Option A, the implied rate of advisers' own choices in BONUS is still significantly larger than in ANTICIPATE.

We can also quantify more exactly the mediating effect that the bonus-induced first recommendations for Option A had on advisers' own choices in BONUS, when they could not foresee the future choices they were required to make. To form a first estimate of this conditional effect, we divide the unconditional effect of the bonus on advisers' own choices in BONUS by its effect on the initial recommendations. This assumes that this channel is the only way that the bonus affected advisers' subsequent own choices. From the unconditional estimates in the first columns of Tables 4 and 5, we get that the increase in O equals 17.3 percentage points, while the increase in R1 is 50.2 percentage points when future actions were not announced beforehand. We then obtain that 34.4% ($\triangleq$ 0.173/0.502) of the advisers whose initial advice shifted towards Option A in BONUS also adjusted their own choices accordingly.

Note that the above estimate is equivalent to the Wald estimate that one obtains in the second stage of a 2SLS-estimation without further controls. In this regression, assignment to the BONUS – as opposed to the NO BONUS – treatment is first used to predict the recommendations for Option A in R1. Then, based on this first stage, the bonus-induced effect of the initial recommendation on advisers' own choices is estimated in the second stage. Column 4 of Table 5, presents these second-stage results, i.e., the local average treatment effect estimates when additional controls are added. It shows that the mediating effect of the bonus' influence on initial recommendations, when this is modeled explicitly through the initial exposure to the bonus. Taking the bonus then corresponds to 38.4 percentage point percentage points in the probability of an adviser later choosing the risky option for himself.

The results look different if the same method is used to investigate whether the initial bonus also affected advisers' subsequent own choices when they could anticipate the upcoming choices that they had to make. Although the first-stage result of the effect of the bonus on the first recommendations are similar between BONUS and ANTICIPATE, they do not spill over on advisers' own choices in the latter treatment. Column 5 of Table 5 shows this. When the same 2SLS-technique is applied to the comparison of NO BONUS and ANTICIPATE, the local average treatment effect of having initially recommended Option A for the bonus on the adviser's subsequent own choices is comparatively small in magnitude and not significantly different from zero. Again, these results are consistent with Prediction 4a. These findings suggest that giving advisers the possibility to evaluate their actions ex-ante leads them to consider the actions as separate decisions and to not factor in image concerns.

Figure 4. Advisers' second recommendations (R2) for Option A over treatments (bars depict standard errors)



5.3 Results for the second recommendations (R2)

For the second recommendation, the decision situation for advisers in NO BONUS is the same as for their first recommendation. Absent image concerns, we therefore expect a similar pattern of recommendations in R2 as in R1 for this treatment. The left bar in Figure 4 supports this notion. Only a small fraction of advisers in NO BONUS recommended Option A – exactly the 3.9% who also recommended this option previously in R1 (see also Table 8 below).

This is very different for the second recommendations in BONUS. Although there is no bonus in R2, the rate of recommendations for Option A is almost six times higher than when there was no previous bonus: 22.9% of the advisers in this treatment recommended Option A, which is a significant increase by 19.0 percentage points relative to NO BONUS ($p = 0.007$). In contrast, the rate of second recommendations for Option A are much lower when there was a bonus but advisers could anticipate this second recommendation. At a level of 6.0%, this rate in ANTICIPATE is significantly lower than in BONUS (Fisher exact test: $p = 0.021$) but not significantly different from the level in NO BONUS (Fisher exact test: $p = 0.678$).

As before, we also conduct a regression analysis by estimating model (9), now with a dummy that indicates whether Option A is recommended in the second recommendation as the dependent variable. Columns 1 and 2 in Table 6 present the results and show that the effect of the BONUS-treatment remains largely unchanged, independently of whether controls are added. The previous findings that this happens only when the adviser's future actions and the bonus' removal were unexpected but not when they could be anticipated are also mirrored in these regression results. Together, these results therefore support Prediction 3 and Prediction 4a.

As for the own choice, we also checked for the mediating effect which the bonus had on the second recommendation through the first recommendation. Column 3 in Table 6 shows that if one includes an indicator for $r_{1,i} = A$ as an additional independent variable, its effect is highly significant while the coefficient of the BONUS-dummy drops and becomes insignificant. Thus, as for the own choice, it is really the bonus' effect on the first recommendation that persistently biases the second one. To measure this effect more precisely, we calculated the share of advisers who re-recommended Option A because they have initially recommended it for the bonus. Column 4 of Table 6 shows this local average treatment effect. This estimate implies that 41.5% of these advisers who initially recommended Option A in BONUS because of the incentive to do so issue the same recommendation again. Column

Table 6. Effect of bonus on advisers' second recommendations in R2

	(1)	(2)	(3)	(4)	(5)
Dependent variable	$r_2 = A$ (second recommendation for Option A)				
BONUS	0.190***	0.193**	0.092		
	(0.067)	(0.078)	(0.080)		
ANTICIPATE	0.021	0.055	-0.058		
	(0.044)	(0.055)	(0.072)		
$r_1 = A$			0.227***		
			(0.080)		
$\widehat{r_1 = A}$ (via BONUS)				0.415***	
				(0.147)	
$\widehat{r_1 = A}$ (via ANTICIPATE)					0.008
					(0.101)
Constant (NO BONUS)	0.039	-0.199	-0.180	-0.331	-0.072
	(0.027)	(0.176)	(0.181)	(0.242)	(0.146)
F-test: BONUS=ANTICIPATE	5.830**	3.430*	3.960**	-	-
Controls	no	yes	yes	yes	yes
Estimation method	OLS	OLS	OLS	2SLS	2SLS
Data used from treatments	B,A,N	B,A,N	B,A,N	B,N	A,N
Observations	149	149	149	99	101

Note: Regression results with robust standard errors in parentheses; significance levels against the (two-sided) null of a zero effect: * p<0.10, ** p<0.05, *** p<0.01. The dependent variable is always a dummy indicating a recommendation for Option A in R2. The main independent variables are dummies indicating whether the adviser was in treatment B(=BONUS) or A(=ANTICIPATE) and whether Option A was recommended in the first recommendation R1; N(=NO BONUS) is the reference category. Additional independent variables control for advisers' age, gender, monthly available budget, region of origin, study degree and field of studies (columns 2–5). Estimates are based on OLS (columns 1–3) or 2SLS where a recommendation for Option A in R1 is instrumented by assignment to treatment B/A and data from treatment A/B is not used (column 4/5, respectively).

5 then shows that, similar as for own choices, this spill-over of giving in to the bonus in R1 does not occur if advisers knew about the upcoming decision situations: The 2SLS-estimate for the effect of initial recommendations, as caused by the bonus, on subsequent recommendations is essentially zero and insignificant in ANTICIPATE. Given that in both, BONUS and ANTICIPATE, the bonus' effect on the initial recommendations was the same, this difference in the local average treatment effect on subsequent recommendations – similar to the difference in this effect for own choices – lends further support for Prediction 4a.

5.4 Further results

There are some additional findings that support our theory and its underlying assumptions. Given our previous results, we expect consistency between advisers' own choices and their first recommendation when there is no conflict of interest. Our results support this notion, as shown in Table 7. It shows the frequency of advisers' own choices over their first recommendations. For NO BONUS, when there is no incentive to bias advice, only the off-diagonal entries are not predicted. They amount to a total of 17.7% of the observations in this treatment; 82.3% of our observations in NO BONUS are therefore consistent with our theory. For BONUS, it predicts that some of those who have previously recommended Option A stick to it in order to avoid a negative self-image. Other advisers who have recommended it but who did not have sufficiently strong image concerns chose their preferred option instead. Accordingly, the theory explains the diagonal entries in the middle three columns of Table 7 plus the off-diagonal ones in the first row of these columns. Again, this leaves only a small fraction, 8.4% of our observations in this treatment, unexplained. A similar picture emerges for observations in ANTICIPATE, presented in the three right-most columns: Again, very few observations, together 6.0%, are not predicted and outside the diagonal and not in the top row. Also note that, in accordance with Prediction 4a and mirroring previous results, the share of advisers who consistently recommend Option A in R1 and choose it for themselves in O in ANTICIPATE is only a third of the corresponding share in BONUS.

The consistency-pattern between advisers' first and second recommendations, displayed in Table 8, is very similar. In NO BONUS, we observe that 17.7% of the second recommendations are inconsistent with the first recommendation, i.e., they are outside the diagonal of Table 8's first three columns. All of them are, however, switches between options C and option B, but not switches to or from the risky Option A. In the BONUS treatment, the results are even stronger. In total, 12.5% of its observations fall outside an explainable pattern, thus are neither on the diagonal nor the first row of Table 8's middle three columns. Again, the picture is similar for treatment ANTICIPATE, where a total of 10.0% of the

Table 7. Frequencies of advisers' own choices conditional on their first recommendation

R1 \ O	NO BONUS			BONUS			ANTICIPATE		
	A	B	C	A	B	C	A	B	C
A	3.9%	0.0%	0.0%	22.9%	8.3%	22.9%	8.0%	20.0%	24.0%
B	2.0%	23.5%	11.8%	0.0%	6.3%	0.0%	0.0%	14.0%	2.0%
C	3.9%	0.0%	54.9%	4.2%	4.2%	31.2%	0.0%	4.0%	28.0%

Table 8. Frequencies of advisers' second recommendations conditional on their first recommendation

R1 \ R2	NO BONUS			BONUS			ANTICIPATE		
	A	B	C	A	B	C	A	B	C
A	3.9%	0.0%	0.0%	18.7%	16.7%	18.8%	4.0%	22.0%	26.0%
B	0.0%	35.3%	2.0%	0.0%	6.2%	0.0%	0.0%	12.0%	4.0%
C	0.0%	15.7%	43.1%	4.2%	8.3%	27.1%	2.0%	4.0%	26.0%

observations is outside the predicted pattern but consistency in recommending Option A is lower than in BONUS. Overall, we find that, in terms of consistency, more than four out of five observations follow a pattern predicted by our theory.

Further evidence comes from our exit questionnaire. Besides questions asking for the subjects' personal characteristics, it also contained a question on advisers' general risk attitudes. More precisely, it asked subjects to indicate on an 11-point Likert-scale "How willing are you to take risk, in general?". This question was not incentivized, but answers to it have previously been shown to correlate with peoples' incentivized choices under risk (see Dohmen et al., 2011). While in NO BONUS, the average response was 5.0 points, it increased by almost one point (39% of the measure's standard deviation) to 5.9 points in BONUS. In a non-parametric test, this difference in the distribution of self-stated risk assessment is marginally statistically significant (Wilcoxon ranksum-test: $p = 0.059$). In contrast, the difference between NO BONUS and ANTICIPATE, where the subjects stated on average 5.3, is just a third of the previous difference and reports in the two conditions are not statistically significant (Wilcoxon ranksum-test: $p = 0.533$).[18]

These results become stronger, both in size and precision, if they are regarded in a regression framework. Columns 1 and 2 of Table 9 present the results from estimating model (9) when the dependent variable is the self-assessed risk-measure without and with control variables. The findings reflect our previous results for O and R2. They therefore suggest that advisers who have previously given in to the bonus can signal that this advice was appropriate, from their point of view, when they consider themselves as more risk-seeking. Also reflecting our previous findings, this persistent effect of the bonus on self-stated risk-tolerance does not occur in ANTICIPATE.

[18] Due to a data glitch in the first two sessions, we had to collect the risk-measure along with the other post-experimental questionnaire data separately for these sessions. When we exclude them all qualitative results remain unchanged. (Also note that our primary data on the recommendations in R1/R2 and on own choices in O were not affected by this data glitch since advisers wrote them on paper.)

Table 9. Effect of bonus on advisers' stated willingness to take risks

	(1)	(2)	(3)	(4)	(5)
Dependent variable		$risk \in \{0, ..., 10\}$ (stated willingness to take risks)			
BONUS	0.914**	1.066**	0.343		
	(0.453)	(0.451)	(0.459)		
ANTICIPATE	0.299	0.151	-0.655		
	(0.472)	(0.530)	(0.518)		
$r_1 = A$			1.615***		
			(0.431)		
$\widehat{r_1 = A}$ (via BONUS)				2.170***	
				(0.825)	
$\widehat{r_1 = A}$ (via ANTICIPATE)					0.308
					(1.138)
Constant (NO BONUS)	4.960***	6.155***	5.822***	7.635***	5.065***
	(0.335)	(1.391)	(1.171)	(1.659)	(1.600)
F-test: BONUS=ANTICIPATE	1.860	3.900**	5.450**	-	-
Controls	no	yes	yes	yes	yes
Estimation method	OLS	OLS	OLS	2SLS	2SLS
Data used from treatments	B,A,N	B,A,N	B,A,N	B,N	A,N
Observations	149	149	149	99	101

Note: Regression results with robust standard errors in parentheses; significance levels against the (two-sided) null of a zero effect: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. The dependent variable is always advisers' stated willingness to take risks (0–10). The main independent variables are dummies indicating whether the adviser was in treatment B(=BONUS) or A(=ANTICIPATE) and whether Option A was recommended in the first recommendation R1; N(=NO BONUS) is the reference category. Additional independent variables control for advisers' age, gender, monthly available budget, region of origin, study degree and field of studies (columns 2–5). Estimates are based on OLS (columns 1–3) or 2SLS where a recommendation for Option A in R1 is instrumented by assignment to treatment B/A and data from treatment A/B is not used (column 4/5, respectively).

Preceding as before to check for the mediating effect of biased first recommendations (and how it differs by what advisers could anticipate) we find in column 3 that the previously positive and significant coefficient for BONUS vanishes when one controls whether the initial recommendation was for Option A. Under the assumption that only the bonus' effect on the initial recommendation caused this shift, we can also compute the bonus' effect on those whose advice it biased. The corresponding estimate in column 4 corresponds to a 2.2-point shift in the self-stated preference for risk for those whose initial advice was biased towards Option A by the bonus and who could not anticipate upcoming actions. The insignificant estimate in column 5 shows that no such mediating effect of the bonus-induced initial recommendation on self-stated risk-assessment occurs when advisers could foresee the upcoming sequence of the actions.

These results show that the bonus' influence on initial recommendations for the risky choice did not only affect advisers' further recommendations and choices when they were unanticipated but also their answers to a more general question with regards to risk.

6 Discussion

Our results show that incentives to bias advice can have a lasting and causal effect on adviser behavior. This is consistent with a psychological mechanism that we propose. This mechanism captures the insight that changing advice after a conflict of interest has disappeared signals that one's initial advice was biased. This mechanism also assumes that issuing biased advice is costly. Only when these costs are sufficiently low relative to image costs, advisers stick to their initial, biased recommendation. In line with this, our estimates imply *partial* consistency. Approximately 40% of the advisers whose initial advice was biased stick to such a recommendation even after the bonus has been removed. We also find a similarly sized, persistent effect of the bonus on the choices of advisers for themselves. This supports the notion that what advisers consider to be appropriate advice is based on their own preferences.[19]

The persistent effect of the initial bonus on advisers' recommendations and their own choices does not appear when advisers know ex-ante that these decision situations will follow the initial recommendation. Also, knowledge of these situations does not decrease the bias in the initial recommendation where the bonus is paid. This result is consistent with our theory based on image concerns but not with anchoring or cue-based theories. It indicates that when advisers can look ahead, they form a plan of action that is not distorted by backward-looking image concerns so that a persistent bias cannot occur.[20]

Additional evidence in this direction comes from subjects' self-stated willingness to take risks. Their responses mirror the results for the bonus' effect on advisers' own choices and repeated recommendations. In particular, having recommended the risky option in BONUS led advisers to state a higher general preference for risk (but not so in ANTICIPATE). Again, this shows that the advisers who were persistently biased by the initial bonus did not act mechanically when they chose for themselves or re-recommended

[19] We did not elicit advisers' belief about their clients' risk preferences, because predicting others' risk preferences is inherently difficult (see Hsee and Weber, 1997; Eckel and Grossman, 2008; Harrison et al., 2013). Even trained advisers who have information about their clients' risk profiles and do not face conflicts of interest often have difficulties to predict their client's risk preferences (see Roth and Voskort, 2014; Kling et al., 2018). Also, if advisers are forced to state ungrounded beliefs about their clients preferences, this could have an effect on advisers' subsequent decisions. That is, it could lead them to act consistently with such stated – but random – beliefs. Such an effect is not what this paper aims to explore.

[20] The results also rule out decreasing absolute risk aversion. Such an explanation would argue that advisers become more risk-seeking in O after earning the 3 GBP. However, this should apply independently of whether they are in BONUS or ANTICIPATE and does therefore not predict the observed differences.

the risky option from the same set of possible options. Rather, this result shows that their initial, biased recommendations have implications that apply to a set of wider but related choices.

Our results can be explained by a single, unifying behavioral mechanism in which backward-looking image concerns are the crucial ingredient. If such image concerns refer to one's self-image, it captures the notion that one constantly learns through one's own actions about the underlying motives of previous actions. Technically, this inferring self is identical to an outside observer who draws such conclusions. In principle, social-image concerns could therefore also be able to capture our findings. However, some features of our experimental design limit such a channel. For example, advisers wrote their recommendations and choices in private and put them in envelopes so that these choices were not exposed. In addition, only one of the first and one of the second recommendations by the advisers in each session were actually shown to the clients. The external consequences of giving biased advice were therefore rather low. Nevertheless, we observe that almost half of the advisers in the treatment with a bonus do not respond to the bonus and that among those who do respond, many acted consistently afterwards. This suggests that the relevant trade-offs occur internally. For these reasons, we interpret our findings as more in line with self-image concerns rather than social-image concerns. However, social-image concerns regarding an actual outside observer follow the same mechanism and could therefore also lead to the same persistent effects. The crucial feature for this would be that such an observer sees the sequence of an adviser's choices and recommendations (e.g., supervisors or regulators who oversee adviser behavior after a new law bans commission-based advice).

7 Conclusion

Our findings have several immediate implications. First, we present evidence that biases in advice-giving can loom longer than the conflict of interest that caused them. Recent policies that ban the causes of conflicts of interest are certainly a correct step towards eventually achieving impartial advice. However, our results show that they should not always be taken as a guarantee that advice becomes immediately impartial, especially when the decision to remove the underlying conflict of interest comes relatively unexpectedly to the people who have been exposed to it.

Second, we also find these persistent effects on advisers' recommendations after incentives to bias them were removed, although they had to choose for themselves before. This observation speaks against the general potential of just allowing advisers to choose for themselves in having a "cleansing effect" on subsequent advice. It also suggests that it can backfire for a company to create conflicts of interest for the advisers who advise external clients when the same persons, for example financial analysts, also affect related decisions within the company.

Third, we show how persistent biases can be deterred. For this, the temporary nature of the bonus and the repeated nature of advice have to be known to the adviser from the beginning. Although this does not diminish the bias in the initial advice, advisers' own choices and repeated recommendations can become unbiased after the conflict of interest is removed. This shows that it can be important for regulators or superiors to inform advisers about upcoming removals of conflicts of interest as soon as possible, even before they become effective. However, this also warrants some caution. Although this seems to be an appealing possibility to prevent persistent biases among early-career advisers who will then perceive, for example, a sales commission as a transitory feature, the effect may be different for more experienced advisers. Experienced advisers might have spent a considerable part of their professional career being exposed to such incentives. The removal of these incentives, even when announced beforehand, may thus be comparatively surprising to them. In addition, image concerns loom larger for experienced advisers as they threaten a considerable part of their professional identities.

Finally, it is important to note that although our findings concern the advice for risky choices, they are not necessarily bound to this specific domain. The crucial feature is that there is no clear-cut right or wrong recommendation so that one can reasonably maintain the image that advice was genuine and unbiased, even though it was not unbiased. Similar effects could therefore also be found for advice on moral, legal or other complex decisions. Investigating these domains would provide interesting avenues for further research.

References

Akerlof, G. A. and W. T. Dickens (1982). The Economic Consequences of Cognitive Dissonance. *American Economic Review 71*(3), 437–447.

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

Benabou, R., A. Falk, and J. Tirole (2018). Narratives, Imperatives and Moral Reasoning. *mimeo*.

Bénabou, R. and J. Tirole (2004). Willpower and Personal Rules. *Journal of Political Economy 112*(4), 848–886.

Bénabou, R. and J. Tirole (2011). Identity, morals, and taboos: Beliefs as assets. *Quarterly Journal of Economics 126*(2), 805–855.

Bénabou, R. and J. Tirole (2016). Bonus Culture: Competitive Pay, Screening, and Multitasking. *Journal of Political Economy 124*(2), 305–370.

Bicchieri, C., E. Dimant, and S. Sonderegger (2019). It's Not A Lie If You Believe It. Lying and Belief Distortion Under Norm-Uncertainty. *mimeo*.

Bodner, R. and D. Prelec (2003). Self-Signaling and Diagnostic Utility in Everyday Decision Making. In I. Brocas and J. D. Carrillo (Eds.), *The Psychology of Economic Decisions*, Volume 1, pp. 105–126. Oxford University Press.

Burks, S. V. and E. L. Krupka (2012). A Multimethod Approach to Identifying Norms and Normative Expectations Within a Corporate Hierarchy: Evidence from the Financial Services Industry. *Management Science 58*(1), 203–217.

Cain, D. M. and A. S. Detsky (2008). Everyone's a Little Bit Biased (Even Physicians). *Journal of the American Medical Association (JAMA) 299*(24), 2893–2895.

Cain, D. M., G. Loewenstein, and D. A. Moore (2005). The Dirt on Coming Clean: Perverse Effects of Disclosing Conflicts of Interest. *Journal of Legal Studies 34*(1), 1–25.

Cain, D. M., G. Loewenstein, and D. A. Moore (2011). When Sunlight Fails to Disinfect: Understanding the Perverse Effects of Disclosing Conflicts of Interest. *Journal of Consumer Research 37*(5), 836–857.

CEA (2015). The Effects of Conflicted Investment Advice on Retirement. *Report by the White House's Council of Economic Advisers*.

Christoffersen, S. E. K., R. Evans, and D. K. Musto (2013). What Do Consumers' Fund Flows Maximize? Evidence from Their Brokers' Incentives. *Journal of Finance 68*(1), 201–235.

Cohn, A., E. Fehr, and M. A. Maréchal (2014). Business culture and dishonesty in the banking industry. *Nature 7529*(516), 86–89.

Cooper, D. J. and J. H. Kagel (2016). Other-Regarding Preferences: A Selective Survey of Experimental Results. In J. H. Kagel and A. E. Roth (Eds.), *The Handbook of Experimental Economics, Volume 2*, Chapter 4, pp. 217–289. Princeton University Press.

Dana, J. and G. Loewenstein (2003). A social science perspective on gifts to physicians from industry. *Journal of the American Medical Association (JAMA) 290*(2), 252–255.

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

Dawes, R. (1990). The potential nonfalsity of the false consens effect. In R. M. Hogarth (Ed.), *Insights in Decision Making: A Tribute to Hillel J. Einhorn*, pp. 179–199. University of Chicago Press.

de Quidt, J., J. Haushofer, and C. Roth (2018). Measuring and Bounding Experimenter Demand. *American Economic Review 108*(11), 3266–3302.

Di Tella, R. D. and R. Pérez-Truglia (2015). Conveniently Upset: Avoiding Altruism by Distorting Beliefs About Others. *American Economic Review 105*(11), 3416–3442.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association 9*(3), 522–550.

Eckel, C. C. and P. J. Grossman (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior and Organization 68*(1), 1–17.

Egan, M., G. Matvos, and A. Seru (2018). The Market for Financial Advice Misconduct. *Journal of Political Economy forthcomin*.

Engelmann, D. and M. Strobel (2000). The false consensus effect disappears if representative information and monetary incentives are given. *Experimental Economics 3*(3), 241–260.

Exley, C. L. (2016). Excusing Selfishness in Charitable Giving: The Role of Risk. *Review of Economic Studies 83*(2), 587–628.

Exley, C. L. and J. B. Kessler (2018). Equity Concerns are Narrowly Framed: Why Money Cannot Buy Time. *mimeo*.

Falk, A. (2017). Facing Yourself: A Note on Self-Image. *HCEO Working Paper 2017-09*.

Falk, A. and F. Zimmermann (2017a). Consistency as a Signal of Skills. *Management Science 63*(7), 2197–2210.

Falk, A. and F. Zimmermann (2017b). Information Processing and Commitment. *Economic Journal*, forthcoming.

Faro, D. and Y. Rottenstreich (2006). Affect, Empathy, and Regressive Mispredictions of Others' Preferences Under Risk. *Management Science 52*(4), 529–541.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.

Festinger, L. and J. M. Carlsmith (1959). Cognitive consequences of forced compliance. *Journal of Abnormal Psychology 58*(2), 203–210.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics 10*(2), 171–178.

Foerster, S., J. T. Linnainmaa, B. T. Melzer, and A. Previtero (2017). Retail Financial Advice: Does One Size Fit All? *Journal of Finance 72*(4), 1441–1482.

Gesche, T. (2016). De-biasing strategic communication? *University of Zurich Department of Economics Working Paper Series 216*(216).

Gino, F., M. I. Norton, and R. A. Weber (2017). Motivated Bayesians : Feeling moral while acting egoistically. *Journal of Economic Perspectives 30*(3), 189–212.

Gneezy, A., U. Gneezy, G. Riener, and L. D. Nelson (2012). Pay-what-you-want, identity, and self-signaling in markets. *Proceedings of the National Academy of Sciences 109*(19), 7236–7240.

Gneezy, U. (2005). Decption: The Role of Consequences. *American Economic Review 95*(1), 384–394.

Gneezy, U., S. Saccardo, M. Serra-Garcia, and R. van Veldhuizen (2016). Motivated Self-Deception, Identity, and Unethical Behavior. *mimeo*.

Gneezy, U., S. Saccardo, and R. van Veldhuizen (2018). Bribery: Behavioral Drivers of Distorted Decision. *Journal of the European Economic Association*, forthcoming.

Grossman, Z. and J. van der Weele (2017). Self-Image and Willful Ignorance in Social Decisions. *Journal of the European Economic Association 15*(1), 173–217.

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

Harrison, G. W., M. I. Lau, E. E. Rutström, and M. Tarazona-Gómez (2013). Preferences over social risk. *Oxford Economic Papers 65*(1), 25–46.

Hoch, S. J. (1987). Perceived consensus and predictive accuracy: The pros and cons of projection. *Journal of Personality and Social Psychology 53*(2), 221–234.

Holt, C. and S. Laury (2002). Risk aversion and incentive effects. *The American Economic Review 92*(5), 1644–1655.

Hsee, C. K. and E. U. Weber (1997). A fundamental prediction error: self-other discrepancies in risk preference. *Journal of Experimental Psychology: General 126*(1), 45–53.

Ifcher, J. and H. Zarghamee (2018). Behavioral Economic Phenomena in Decision-Making for Others. *IZA Discussion Paper 11946*.

Inderst, R., K. Khalmetski, and A. Ockenfels (2018). Sharing Guilt: How Better Access to Information May Backfire. *Management Science*, forthcoming.

Inderst, R. and M. Ottaviani (2012). Competition through Commissions and Kickbacks. *American Economic Review 102*(2), 780–809.

Kerschbamer, R. and M. Sutter (2017). The economics of credence goods - A survey of recent lab and field experiments. *CESifo Economic Studies 63*(1), 1–23.

Kling, L., C. König-Kersting, and S. T. Trautmann (2018). Investing for Others : Principals ' vs . Agents ' Preferences. *mimeo*.

Koch, C. and C. Schmidt (2010). Disclosing conflicts of interest - Do experience and reputation matter? *Accounting, Organizations and Society 35*(1), 95–107.

Konow, J. (2000). Fair Shares : Accountability and Cognitive Dissonance in Allocation Decisions Fair Shares. *American Economic Review 90*(4), 1072–1091.

Kunda, Z. (1992). Can Dissonance Theory Do It All? *Psychological Inquiry 4*(3), 337–339.

Li, M. and K. Madarasz (2008). When mandatory disclosure hurts: Expert advice and conflicting interests. *Journal of Economic Theory 139*, 47–74.

Linnainmaa, J. T., B. Melzer, and A. Previtero (2018). The Misguided Beliefs of Financial Advisors. *Journal of Finance*, forthcoming.

Loewenstein, G., S. Issacharoff, C. Camerer, and L. Babcock (1993). Self-Serving Assessments of Fairness and Pretrial Bargaining. *Journal of Legal Studies 22*(1), 135–159.

Loewenstein, G., C. R. Sunstein, and R. Golman (2014). Disclosure: Psychology Changes Everything. *Annual Review of Economics 6*, 391–419.

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

Malmendier, U. and D. Shanthikumar (2014). Do security analysts speak in two tongues? *Review of Financial Studies 27*(5), 1287–1322.

Marks, G. and N. Miller (1987). Ten years of research on the false-consensus effect: An empirical and theoretical review. *Psychological Bulletin 102*(1), 72–90.

Mazar, N., O. Amir, and D. Ariely (2008, dec). The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research 45*(6), 633–644.

Mijović-Prelec, D. and D. Prelec (2010). Self-deception as self-signalling: a model and experimental evidence. *Philosophical Transactions of the Royal Society B 365*, 227–40.

Mullainathan, S., M. Noeth, and A. Schoar (2012). The Market For Financial Advice. An Audit Study. *NBER Working Paper Series 17929*.

Mullen, B., J. L. Atkins, D. S. Champion, C. Edwards, D. Hardy, J. E. Story, and M. Vanderklok (1985). The false consensus effect: A meta-analysis of 115 hypothesis tests. *Journal of Experimental Social Psychology 21*(3), 262–283.

Rabin, M. (1994). Cognitive dissonance and social change. *Journal of Economic Behavior & Organization 23*, 177–194.

Rode, J. (2010). Truth and trust in communication: Experiments on the effect of a competitive context. *Games and Economic Behavior 68*(1), 325–338.

Ross, L., D. Greene, and P. House (1977). The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology 13*(3), 279–301.

Roth, B. and A. Voskort (2014). Stereotypes and false consensus: How financial professionals predict risk preferences. *Journal of Economic Behavior and Organization 107*, 553–565.

Schwardmann, P. and J. van der Weele (2016). Decepetion and Self-deception. *mimeo*.

Sutter, M. (2009). Deception through Telling the Truth ?! Experimental Evidence from Individuals and Teams Author. *Economic Journal 119*(534), 47–60.

Trivers, R. (2011). *The folly of fools: The logic of deceit and self-deception in human life*. Basic Books.

Tversky, A. and D. Kahneman (1974). Judgment under Uncertainty: Heuristics and Biases. *Science 185*(4157), 1124–1131.

Zingales, L. (2015). Presidential Address: Does Finance Benefit Society? *Journal of Finance 70*(4), 1327–1363.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics 13*(1), 75–98.

For Online Publication

Persistent Bias in Advice-Giving

Online Appendix

Contents:

A: A signaling model of persistent biases in advice-giving

B: Proofs for Appendix A

C: Additional data

D: Experimental instructions

Contact: chenzq926@gmail.com and tobias.gesche.job@gmail.com

Appendix A: A model of persistent biases in advice-giving (full version)

In the following, we set up a formal model which demonstrates how advisers can be affected by conflicts of interest, even after they have been removed. The key assumptions underlying it reflect those described in Section 2. In the model, we establish Corollary 1 through 4 which are analogous to the respective predictions in the main text.

A1. Model setup

We consider an adviser who advises a client on which element out of a discrete, finite set of possible choices $\mathcal{C}$ to take. The adviser may also have to choose for himself from this set. He gets a bonus payment $b \geq 0$ if he recommends an option from the set $\mathcal{B} \subset \mathcal{C}$. Hence, when $b > 0$, the adviser is subject to a bias towards recommending a choice from $\mathcal{B}$.[1] We denote by $\mathcal{N}$ the subset of choices with no bonus, i.e. $\mathcal{N} = \mathcal{C}/\mathcal{B}$. Both, $\mathcal{B}$ and $\mathcal{N}$ are assumed to be non-empty. Three factors influence an adviser's actions: 1) his own (expected) pecuniary payoff, 2) costs of giving inappropriate advice, and 3) image concerns of being perceived of having given biased advice. We will explain them in detail below:

Adviser's pecuniary payoff and personal preferences: We assume that pecuniary payoffs map into the adviser's utility via the strictly increasing vNM-utility function $u : \mathbb{R} \rightarrow \mathbb{R}$ with $u(0) = 0$. Let $v(a)$ denote the corresponding pecuniary payoff which an adviser gets from action $a \in \mathcal{C}$. For a choice $o \in \mathcal{C}$ made for the adviser himself, $v(o)$ represents his certainty equivalent of the option. For a recommendation, $v(r) = b \cdot \mathbb{1}[r \in \mathcal{B}]$ where $\mathbb{1}[\cdot]$ denotes the indicator function which takes a value of one if the statement in the bracket is true. The own choice o which an adviser optimally chooses for himself from a, possibly restricted, subset $\mathcal{X} \subseteq \mathcal{C}$ is denoted by $o^*_{\mathcal{X}} \equiv \arg\max_{o \in \mathcal{X}}\{u(v(o))\}$. To save on notation we assume w.l.o.g. that $o^*_{\mathcal{X}}$ is a singleton for each $\mathcal{X} \subseteq \mathcal{C}$. The subscript is omitted when the choice-set is unrestricted, i.e., $o^* = o^*_{\mathcal{C}}$. The share of advisers for whom $o^* \in \mathcal{X} \subseteq \mathcal{C}$, i.e., whose unconstrained optimum lies in $\mathcal{X}$ will be denoted with $\alpha_{\mathcal{X}}$ (thus, for such advisers $o^* = o^*_{\mathcal{X}}$ holds).

Costs of giving inappropriate advice: Each adviser has a single choice which he consider appropriate to recommend. This "appropriate recommendation" is denoted by $r^* \in \mathcal{C}$. We denote with $\beta_{\mathcal{X}}$ the share of advisers for whom $r^* \in \mathcal{X} \subseteq \mathcal{C}$, i.e., those who think that the option they consider appropriate is in $\mathcal{X}$. In the following, we will consider two prominent possibilities of how r^* is determined:

- *Projected appropriate recommendations* ($r^* = o^*$)*:* As mentioned in the main text, there is ample evidence that people project their own preferences onto others, e.g. clients. Equivalently, they might follow a general rule which stipulates, for themselves and others equally, what ought to be chosen. This means that $r^* = o^*$ holds and therefore $\alpha_{\mathcal{X}} = \beta_{\mathcal{X}}$ for each $\mathcal{X} \subseteq \mathcal{C}$.
- *Independent appropriate recommendations* ($r^* \perp\!\!\!\perp o^*$)*:* Alternatively, advisers may base appropriate recommendations on criteria which are unrelated to their own preferences. For example, they can hold a belief about the client's preferences which then stipulates which choice would suit the client best. Importantly,

[1] Note that this is equivalent to a punishment $p = -b$ he has to pay if he does not recommend an option from $\mathcal{B}$.

such a belief can be motivated and instrumental in helping the adviser to recommend a choice he would not prefer for himself. In such a setting, o^* and r^* and their respective distributions are independent.

Giving inappropriate advice creates costs for the adviser. In the context of our experiment these costs are psychological but they could also be expected legal costs or both. They are captured by the dis-utility $\kappa \geq 0$ which an adviser experiences if he recommends an option $r \in \mathcal{C}$ when $r \neq r^*$.

Image costs of being perceived as biased: In addition to the immediate costs of not recommending what is considered appropriate, we also allow for costs of being *perceived* ex-post of having recommended in such a manner. More precisely, we assume that the adviser suffers dis-utility $\lambda \geq 0$ to the degree that he (or someone else who observes his actions) learns that previous advice was biased, i.e., that a previous recommendation r did not correspond to the adviser's appropriate action r^*. This "degree", which weighs these costs, corresponds to the posterior probability that, given an adviser's prior and current actions, previous advice was biased. The observer who makes such an inference observes the adviser's actions but not r^*, the choice which the adviser considers to be appropriate advice. Given our setup and findings, we interpret such image concerns as self-image concerns. This corresponds to a dual-self model, similar to Bodner and Prelec (2003) or Bénabou and Tirole (2011): One self is a standard economic agent who trades off the benefits and costs of any action and knows whether the adviser gave inappropriate advice or not, e.g. via what Bodner and Prelec (2003) call "gut-feeling". The other self does not know this and is modeled as an outside observer who only sees an adviser's actions. Thus, *social* image concerns regarding an actual outside observer follow the same model.

Payoff and utility function: Given an adviser's history h_a of choices and recommendations prior to action a and the choice r^* he considers appropriate, his overall utility can then be written as follows:

$$\begin{aligned} U(a \mid h_a, r^*) = u(v(a)) - \kappa \cdot \mathbb{1}[a \neq r^* \textit{ and } a \textit{ is a recommendation}] \\ - \lambda \cdot \Pr[\textit{previous advice was biased} \mid h_a, a] \end{aligned} \tag{1}$$

The utility function (1) is a generalized version of the utility function (1) in Section 4. In other words, function (1) is an application of utility (1) to our specific experimental setup. The same as in Section 4, the first term denotes the adviser's utility from pecuniary payoffs. The second term denotes the costs of recommending something which is not considered appropriate advice. The third term is the expected image costs of being perceived as biased. Therefore, the adviser chooses a such that $U(a \mid h_a, r^*)$ is maximized, given that $\Pr[\textit{previous advice was biased} \mid h_a, a]$ is updated via Bayes' rule under knowledge of the adviser's current action a and the history h_a prior to this action. We focus on pure strategies. As a tie-breaking rule we make the (natural) assumption that if an adviser is indifferent between multiple choices for himself which includes o^*, his preferred choice, he chooses o^*. Similarly, if he is indifferent between recommending different choices of which one is r^*, the choice he considers appropriate, he recommends r^*.

Heterogeneity and information structure: Advisers' moral and image costs are heterogeneous. We denote the corresponding joint distribution via its c.d.f. $J(x, y) = \Pr[\kappa \leq x, \lambda \leq y]$. To save considerably on notation, we assume that κ and λ are independent of o^* and r^*.[2] Note that this does not prevent λ and κ to be correlated. We can then state the following:

Lemma 1. *Suppose the joint distribution of* (κ, λ) *is absolutely continuous and the associated p.d.f. has full support over* $\mathbb{R}_0^+ \times \mathbb{R}_0^+$. *Then, the following holds:*

a) The marginal c.d.f.s $K(x) = \Pr[\kappa \leq x]$ *and* $\Lambda(y) = \Pr[\lambda \leq y]$ *are strictly increasing for every* $x, y \geq 0$.

b) The conditional marginal c.d.f. $\Lambda(y \mid x) = \Pr[\lambda \leq y | \kappa \leq x]$ *is strictly increasing in* y *for every* $y \geq 0$ *and for any* $x > 0$.

c) The conditional c.d.f. for the distribution of the ratio $(\kappa/\lambda \mid \lambda > 0)$, *given by* $R(z \mid x, y) = \Pr[\kappa/\lambda \leq z \mid \kappa \leq x, \lambda \geq y]$, *exists and is non-decreasing in* z *for every* $x, y > 0$.

Proof: see Appendix B.

In the following, we assume that the above assumptions and, therefore, Lemma 1 hold. We also assume that the joint distribution J, together with the families of distributions $\{\alpha_{\mathcal{X}}\}_{\mathcal{X} \subseteq \mathcal{C}}$ and $\{\beta_{\mathcal{X}}\}_{\mathcal{X} \subseteq \mathcal{C}}$ which describe the distribution of advisers' preferences and what they consider appropriate recommendations, are common knowledge. To make things interesting, we also assume that some, but not all, advisers consider an option appropriate which would not earn them the bonus, i.e., $\beta_{\mathcal{N}} \in (0, 1)$. While an adviser knows his individual values of (κ, λ, r^*), the observer or the observing self does not know this preference vector. However, the distributions of the vector's elements are, as they are described above, common knowledge.

Solution concepts: Our analysis of how a one-off incentive can lead to a persistent bias in advice-giving follows closely with our experimental design. For this, we consider a situation in which an adviser first has to issue a recommendation $r_1 \in \mathcal{C}$ for which he can earn a bonus b, and then he make a choice for himself among the same set of choices and a second recommendation $r_2 \in \mathcal{C}$ to another client for which no bonus can be earned.

In our treatments where advisers could not anticipate the stages O and R2 which followed R1, each of these stages is effectively a static game with a given history h_a of actions prior to $a \in \{r_1, o, r_2\}$. We make predictions for these treatments by solving for Bayesian Nash Equilibrium. That is, advisers choose their choices or recommendation from $\mathcal{C}$ such that it maximizes their over utility $U(a \mid h_a, r^*)$ as defined in (1). In particular, they take into account the implications their action a has through updating their belief $\Pr[$*previous advice was biased* $\mid h_a, a]$, given the respective prior history h_a (where $h_a = \emptyset$ if $a = r_1$, $h_a = r_1$ if $a = o$, and $h_a = (r_1, o)$ if $a = r_2$), using Bayes' rule.

[2] With such correlation all our results would remain valid if the joint distributions of (κ, λ), conditional on a preferred own choice o^* and appropriate choice r^* are increasing. For example, full support for $J_c(x, y) \equiv \Pr[\kappa \leq x, \lambda \leq y \mid o^* = c]$ and $\tilde{J}_c(x, y) \equiv \Pr[\kappa \leq x, \lambda \leq y \mid r^* = c]$ for all $c \in \mathcal{C}$ would be such a sufficient (but not necessary) condition.

A2. Analysis of BONUS and NO BONUS

Our experiment resembles this setting with $\mathcal{C} = \{A, B, C\}$ and $\mathcal{B} = \{A\}$ and where in the BONUS-treatment $b = 3$ GBP holds. It also includes a counter-factual where no incentive to bias advice is ever present, the NO BONUS-treatment with $b = 0$ GBP. In addition to repeated advice-giving, our experiment also features a stage where, after having made the first recommendation but before the second, the adviser has to make an own choice $o \in \mathcal{C}$ for himself. For this own choice, no bonus can be earned either. This allows us to separate whether advisers form motivated beliefs or whether they tie advice to own preferences which prevents such self-serving beliefs. However, as will become clear, the main result regarding the persistent bias in advice-giving is independent of whether there is an own choice or not.

Advisers' behavior is analyzed step by step, in the order as subjects acted in the experiment: We start with the first recommendation (R1), then treat the own choice (O), and finally cover the second recommendation (R2). In each step we contrast behavior when there was an initial conflict of interest (BONUS) with behavior when there was no such conflict (NO BONUS).

First recommendation R1

R1 – NO BONUS: Here, the adviser's action a is a recommendation denoted by $a = r_1$. There is no prior advice and therefore, image concerns do not matter. Using that $v(r_1) = 0$ because $b = 0$, (1) becomes $U(r_1 \mid h_{r_1}, r^*) = -\kappa \cdot \mathbb{1}[r_1 \neq r^*]$ where $h_{r_1} = \emptyset$. Accordingly, advisers recommend $r_1 = r^*$ and the share of advisers who recommend an option from $\mathcal{B}$ is given by $\beta_{\mathcal{B}}$.

R1 – BONUS: Recommending an option from $\mathcal{B}$ now yields the bonus, captured by $v(r_1) = b \cdot \mathbb{1}[r_1 \in \mathcal{B}]$ with $b > 0$. There is no previous advice, so that image concerns do not matter. Thus, (1) becomes $U(r_1 \mid h_{r_1}, r^*) = u(b \cdot \mathbb{1}[r_1 \in \mathcal{B}]) - \kappa \cdot \mathbb{1}[r_1 \neq r^*]$ where $h_{r_1} = \emptyset$. For the share $\beta_{\mathcal{B}}$ of advisers who have $r^* \in \mathcal{B}$, recommending $r_1 = r^*$ is then clearly optimal – they get rewarded for what they would have recommended anyway. However, for a share $1 - \beta_{\mathcal{B}}$ of advisers, $r^* \in \mathcal{N}$, holds. They face a trade-off between recommending an option from $\mathcal{B}$ even though they do not consider it appropriate and being impartial. Recommending an option from $\mathcal{B}$ yields them pecuniary utility $u(b)$ but causes costs κ of giving inappropriate advice. Being impartial by recommending $r_1 = r^* \in \mathcal{N}$ does not create such costs but no bonus is earned either. Accordingly, those with costs κ lower than $u(b)$ give biased advice; their population share is given by $(1 - \beta_{\mathcal{B}}) \cdot K(u(b)) > 0$. It follows that advisers' behavior in the BONUS-treatment corresponds to three different behavioral types, denoted by $\theta \in \{1, 2, 3\}$ and determined by their values for κ, λ, and r^*:

- $\theta = 1$ – unbiased advisers who recommend an option which earns them a bonus because they truly think that it is appropriate for the client ($r_1 = r^* \in \mathcal{B}$). Their population share is $\phi_1 \equiv \beta_{\mathcal{B}}$.
- $\theta = 2$ – unbiased advisers who recommend an option which does not earn them a bonus because they think that this option is appropriate ($r_1 = r^* \in \mathcal{N}$) and who are not biased by the bonus because their κ is sufficiently high. Their population share is $\phi_2 \equiv (1 - \beta_{\mathcal{B}}) \cdot (1 - K(u(b))) > 0$.

- $\theta = 3$ – biased advisers who recommend an option which earns them a bonus even though they do not think that it is appropriate to do so ($r_1 \in \mathcal{B}$ but $r_1 \neq r^* \in \mathcal{N}$). They are biased by the bonus because their κ is low enough. Their population share is $\phi_3 \equiv (1 - \beta_{\mathcal{B}}) \cdot K(u(b)) > 0$.

Note that by letting $b = 0$, the above also applies to the NO BONUS-treatment. In this case, only share ϕ_1 recommends an option from $\mathcal{B}$ as there are no type-3-advisers. Also note that from the above, $\Pr[$*previous advice was biased* $\mid h_a, a] = \Pr[\theta = 3 \mid h_a, a]$ holds. We then get the following corollary, leading to Prediction 1 in the main text:

Corollary 1. *The share of advisers who recommend a choice from $\mathcal{B}$ in BONUS is given by $\phi_1 + \phi_3$ and is larger than the share ϕ_1 of advisers who recommend such a choice in NO BONUS.*

Own choice O

O – NO BONUS: The action a is now the adviser's choice for himself and denoted by $a = o$. Accordingly, its corresponding pecuniary value $v(o)$ is his certainty equivalent of the choice o. As this is not an advice to a client, the costs κ of giving inappropriate advice do not matter. Without a bonus, only type-1 and type-2-advisers exist so that there are also no concerns of being perceived as biased. Therefore, (1) becomes $U(o \mid h_o, r^*) = u(v(o))$ which is maximized by an adviser's preferred own choice o^*. Thus, the share of advisers choosing an option from $\mathcal{B}$ is given by $\alpha_{\mathcal{B}}$.

O – BONUS: As there were biased advisers in the previous recommendation R1 (the type-3-advisers) image costs of being perceived as them matter. Given the prior history $h_o = r_1$ and the current choice for oneself $a = o$, the adviser's objective function (1) becomes $U(o \mid h_o, r^*) = u(v(o)) - \lambda \cdot \Pr[\theta = 3 \mid h_o, o]$. It therefore matters whether o has diagnostic value:

- *Projected appropriate recommendations:* This means that $r^* = o^*$. An intuitive implication is then that, were it not for the bonus, advisers should choose what they have recommended. Type-3-advisers who have previously recommended $r_1 \neq o^* = r^*$ would be put on the spot: By choosing $o = o^* \neq r_1$ they would reveal themselves as type-3-advisers with $r_1 \neq r^*$ because type-1 and type-2-advisers choose $o = o^* = r_1 = r^*$. Type-3-advisers would then suffer full dis-utility λ for the benefit of choosing their own preferred choice. Alternatively, type-3-advisers could pool with type-1-advisers by choosing $o = r_1 \in \mathcal{B}$. By this, they would lower the weight on the image costs of being perceived as biased but incur costs of not choosing what they actually prefer because for them, $o^* \in \mathcal{N}$ holds. The following proposition shows that such behavior is indeed the unique equilibrium in this situation and that some, but not all, type-3-advisers choose a non-preferred choice for themselves to pool with type-1-advisers:

  Proposition 1. *When $b > 0$ and $o^* = r^*$, there is a unique equilibrium in which all advisers of type $\theta \in \{1, 2\}$ and a share $\pi_o^* \in (0, 1)$ of advisers of type $\theta = 3$ choose $o = r_1$. The remaining share of type-3-advisers chooses $o \neq r_1$.*

  Proof: see Appendix B.

- *Independent appropriate recommendations:* Having a (possibly self-serving) belief about what constitutes appropriate advice prevents the above-described pressure on type-3-advisers. Such a belief prevents any inference via o on whether $r_1 = r^*$ holds, i.e., $\Pr[\theta = 3 \mid h_o, o]$ is invariant to o. The choice $o \in \mathcal{C}$ which maximizes $U(o \mid h_o, r^*) = u(v(o)) - \lambda \cdot \Pr[\theta = 3 \mid h_o, o]$ is then the one which maximizes its first element, given by o^*.

The result below then follows directly from the above and describes the situation for own choice O across the two environments with and without a bonus:

Corollary 2. *If own choices and appropriate recommendations are identical (projected appropriate recommendations), the share of advisers choosing $o \in \mathcal{B}$ in BONUS is given by $\phi_1 + \pi_o^* \phi_3$ with $\pi_o^* \in (0, 1)$ and is larger than ϕ_1, the share of advisers who make such a choice in NO BONUS. If own choices and appropriate recommendations are independent (independent appropriate recommendations), the share of advisers choosing $o \in \mathcal{B}$ is given by ϕ_1 in both, BONUS and NO BONUS.*

Second recommendation R2

R2 – NO BONUS: Here, a is another recommendation, denoted by $a = r_2$. As with the first recommendation in NO BONUS, there is no payment involved, thus $v(r_2) = 0$. Also, there are no type-3-advisers in this condition so that image concerns do not matter. Since r_2 is a recommendation, costs of giving inappropriate advice matter and (1) becomes $U(r_2 \mid h_{r_2}, r^*) = -\kappa \cdot \mathbb{1}[r_2 \neq r^*]$ where $h_{r_2} = (r_1, o)$. Recommending $r_2 = r^*$ maximizes this expression so that share ϕ_1 of advisers recommend an option from $\mathcal{B}$.

R2 – BONUS: In this condition, the initial bonus has been removed so that $v(r_2) = 0$ holds, as in NO BONUS. However, as there was a bonus in the first recommendation, there is a positive mass of type-3-advisers and image concerns matter, in addition to the costs of giving inappropriate advice. The advisers' utility (1) then becomes $U(r_2 \mid h_{r_2}, r^*) = -\kappa \cdot \mathbb{1}[r_2 \neq r^*] - \lambda \cdot \Pr[\theta = 3 \mid h_{r_2}, r_2]$ where $h_{r_2} = (r_1, o)$.

Similar to own choices under projected appropriate recommendations, this puts type-3-advisers on the spot again: If advice in the initial recommendation was unbiased, this advice should be issued again as the presence of a bonus should not have affected the initial recommendation r_1. Accordingly, type-1 and type-2-advisers should recommend $r_2 = r_1$. Type-3-advisers who have not yet revealed themselves as such could then re-recommend $r_2 = r_1 \in \mathcal{B}$ in order to pool with type-1-advisers which discounts their image costs λ. However, since for them $r^* \in \mathcal{N}$ holds, they would suffer costs κ of issuing biased advice (again). If they want to prevent these costs and recommend $r_2 = r^*$, this means that $r_2 \neq r_1$ so that they would reveal themselves as type-3-advisers and suffer full image costs λ. They do so if λ is small, relative to the costs κ of re-issuing biased advice.

The above reasoning applies to type-3-advisers who have not yet revealed themselves. This is always the case when appropriate recommendations are independent. In contrast, when appropriate recommendations are projected, some type-3-advisers have already revealed themselves by choosing $o \neq r_1$. For them, there is no point of trying to pool with type-1s. However, as Proposition 1 shows, there is a non-zero share π_o^* of type-3-advisers who have not yet revealed themselves and to whom the preceding reasoning applies. Proposition 2 summarizes this and proves that the partial pooling described above is the unique equilibrium outcome.

Proposition 2. *When $b > 0$, there is a unique equilibrium in which all advisers of type $\theta \in \{1, 2\}$ and share $\psi \cdot \pi^*_{r_2}$ of advisers of type $\theta = 3$ choose $r_2 = r_1$, the other advisers with $\theta = 3$ recommend $r_2 \neq r_1$. For this, it always holds that $\pi^*_{r_2} \in (0, 1]$. If $o^* = r^*$ (projected appropriate recommendations), then $\psi = \pi^*_o$. If o^* and r^* are independent (independent appropriate recommendations), then $\psi = 1$.*

Proof: see Appendix B.

Corollary 3. *The share of advisers re-recommending a choice from $\mathcal{B}$ in BONUS is given by $\phi_1 + \psi\pi^*_{r_2}\phi_3$ and is larger than ϕ_1, the share of advisers who recommend such a choice in NO BONUS.*

A3. Analysis of ANTICIPATE

We now consider a situation where the adviser can anticipate upcoming actions and the bonus' one-off nature. Forming a plan over current and future decision situation can therefore be understood as picking a vector $(r_1, o, r_2) \in \{\mathcal{B}, \mathcal{N}\} \times \{\mathcal{B}, \mathcal{N}\} \times \{\mathcal{B}, \mathcal{N}\}$. We adapt a short-hand notation for the $2 \times 2 \times 2 = 8$ possible combinations of adviser actions over these subsets. For example, a sequence of actions $(r_1, o, r_2) \in \mathcal{B} \times \mathcal{N} \times \mathcal{B}$ is written simply as $\mathcal{BNB}$ and means that the first and second recommendation r_1 and r_2 are for an option from $\mathcal{B}$ whereas the own choice o is from $\mathcal{N}$.

When advisers form a plan for R1, O, and R2, there are two possibilities: The first is that they ex-ante consider the decisions for R1, O, and R2 as a one-stage decision with three elements (r_1, o, r_2). In this case, $h_o = h_{r_2} = \emptyset$ holds because there is no prior history for O and R2 since these decisions are made simultaneously with R1. The second is that they consider the decisions for R1, O, and R2 as one sequence of decisions with three interacting elements. In that case, we have $h_o = r_1$ and $h_{r_2} = (r_1, o)$ as in our analysis for BONUS. In either case, advisers in ANTICIPATE are aware that they will have to make three decisions (r_1, o, r_2). Their ex-ante anticipated utility therefore corresponds to the sum of utilities for the initial recommendation r_1, the own choice o, and the second recommendation r_2. These were analyzed sequentially in BONUS and are now analyzed altogether, as given by the following expression. In it, each line corresponds to the effect on the payoffs in R1, O, and R2, respectively:

$$\begin{aligned}\sum_{a \in \{r_1, o, r_2\}} U(a \mid h_a, r^*) \quad = \quad & u(v(r_1)) - \kappa \cdot \mathbb{1}[r_1 \neq r^*] \\ & + u(v(o)) - \lambda \cdot \Pr[\theta = 3 \mid h_o, o] \\ & - \kappa \cdot \mathbb{1}[r_2 \neq r^*] - \lambda \cdot \Pr[\theta = 3 \mid h_{r_2}, r_2] \end{aligned} \tag{2}$$

Advisers consider choices as one-stage decisions

If advisers consider the three decisions as a single, one-stage decision, then $h_o = h_{r_2} = \emptyset$ and thus, $\Pr[\theta = 3 \mid h_o, o] = \Pr[\theta = 3 \mid h_{r_2}, r_2] = 0$ holds for the above utility function. Hence, advisers do not factor in image concerns in this case. We then start the analysis of this situation by considering the decision of advisers with $r^* \in \mathcal{B}$. It is straightforward to check that $\mathcal{BBB}$ dominates all alternatives since the corresponding payoff is $u(b) + u(v(o^*))$ whereas all other choices involve giving inappropriate advice or making a suboptimal choices.

For advisers with $r^* \in \mathcal{N}$ who do not factor in image concerns in their ~~initial~~ decisions, the relevant payoffs change to those listed below:

$$\mathcal{BBB}: \quad u(b) + u(v(o^*_{\mathcal{B}})) - 2\kappa \tag{3}$$

$$\mathcal{BBN}: \quad u(b) + u(v(o^*_{\mathcal{B}})) - \kappa \tag{4}$$

$$\mathcal{BNB}: \quad u(b) + u(v(o^*)) - 2\kappa \tag{5}$$

$$\mathcal{BNN}: \quad u(b) + u(v(o^*)) - \kappa \tag{6}$$

$$\mathcal{NNN}: \quad u(v(o^*)) \tag{7}$$

$$\mathcal{NBN}: \quad u(v(o^*_{\mathcal{B}})) \tag{8}$$

$$\mathcal{NBB}: \quad u(v(o^*_{\mathcal{B}})) - \kappa \tag{9}$$

$$\mathcal{NNB}: \quad u(v(o^*)) - \kappa \tag{10}$$

Simple comparisons show that (6) dominates (3), (4), and (5) while (7) dominates (8), (9), and (10). Hence, these advisers would either choose $\mathcal{BNN}$ or $\mathcal{NNN}$. They choose the former over the latter whenever $u(b) + u(v(o^*)) - \kappa > u(v(o^*))$, leading to the following result:

Proposition 3. *Suppose advisers consider choices in ANTICIPATE as one-stage decisions. All advisers with $r^* \in \mathcal{B}$ choose $\mathcal{BBB}$. The share of advisers with $r^* \in \mathcal{N}$ choose $\mathcal{BNN}$ is given by $K(u(b))$. All other advisers with $r^* \in \mathcal{N}$ choose $\mathcal{NNN}$.*

Proof. See Appendix B. □

Recall from the previous subsection that when there is a bonus and actions could not be anticipated, all type-1-advisers recommend $r_1 \in \mathcal{B}$. Also, a share $K(u(b))$ of advisers with $r^* \in \mathcal{N}$ recommend such an option. For the case of no bonus, we got that only type-1-advisers with $r^* \in \mathcal{B}$ choose $o \in \mathcal{B}$ and recommend $r_2 \in \mathcal{B}$. The above proposition shows that when anticipation was possible, the behavior for type-1-advisers is the same. This implies the following corollary, leading to Prediction 4a:

Corollary 4a. *If advisers consider choices in ANTICIPATE as one-stage decisions,*

a) the share of advisers who recommend $r_1 \in \mathcal{B}$ in ANTICIPATE is given by $\beta_{\mathcal{B}} + (1 - \beta_{\mathcal{B}})K(u(b))$ and equals the corresponding share in BONUS,

b) the share of advisers who choose $o \in \mathcal{B}$ in ANTICIPATE is given by $\beta_{\mathcal{B}}$ and equals the corresponding share in NO BONUS (and is therefore lower than in BONUS), *c) the share of advisers who recommend $r_2 \in \mathcal{B}$ in ANTICIPATE is given by $\beta_{\mathcal{B}}$ and equals the corresponding share in NO BONUS (and is therefore lower than in BONUS).*

Advisers consider choices as multi-stage decisions

Suppose advisers consider decisions as multi-stage and therefore anticipate the potential future image costs. This means that they find themselves in a standard dynamic game with three stages. Then, it is true in advisers' utility function (2) that $h_o = r_1, h_{r_2} = (r_1, o)$. For this case with the dynamic nature, the static solution concept of BNE we used for the previous analysis no longer applies. we make our predictions through solving for the following Perfect Bayesian Equilibrium of the model:

Definition 1. *A Perfect Bayesian Equilibrium (PBE) of the ANTICIPATE treatment is a vector of actions and beliefs* $((r_1, o, r_2), \Pr[r_1$ *was biased* $\mid h_a, a])$ *such that*

i) (r_1, o, r_2) *maximizes payoff function* $\sum_{a\in\{r_1,o,r_2\}} U(a|h_a, r^*)$ *given* $\Pr[r_1$ *was biased* $\mid h_a, a]$*;*

ii) $\Pr[r_1$ *was biased* $\mid h_a, a]$ *is calculated using Bayes' rule.*

For ease of notation, we define the following abbreviations for the posterior that an adviser is type-3 ($\theta = 3$) conditional on his actions o, r_2 and corresponding histories $h_o = r_1, h_{r_2} = (r_1, o)$, respectively:

$$\begin{aligned} P_{M_1M_2} &\equiv \Pr[\theta = 3 \mid r_1 \in M_1, o \in M_2] \\ P_{M_1M_2M_3} &\equiv \Pr[\theta = 3 \mid (r_1 \in M_1, o \in M_2), r_2 \in M_3] \end{aligned}$$

where $M_1, M_2, M_3 \in \{\mathcal{B}, \mathcal{N}\}$. For example, $P_{\mathcal{BNN}} = \Pr[\theta = 3 \mid (r_1 \in \mathcal{B}, o \in \mathcal{N}), r_2 \in \mathcal{N}]$.

To make things clear, let us first write down the payoffs of advisers for all eight choices. We will make use again of our notation $o^*_{\mathcal{X}}$ for the optimal choice from a (possibly restricted) subset $\mathcal{X} \subseteq \mathcal{C}$. As before, we omit the subscript when the choice set is unrestricted, thus $o^* = o^*_{\mathcal{C}}$. We also assume that $o^* = r^*$ holds. This corresponds to the case that appropriate recommendations are projected (see above, in particular Prediction 2a and the second part of Corollary 2). This is done because our results in O actually support this notion (see subsection 5.2) and because it considerably simplifies the exposition by limiting it to relevant cases.[3]

[3] One can also model the case when own choices and appropriate action are independent. In such a setting, the share of advisers who choose $o \in \mathcal{B}$ in ANTICIPATE is the same as the share in NO BONUS. This is because own choices do not have image implications in this case. Hence, advisers choose the option which maximizes their expected utility. However, the second recommendation still has image implications. Advisers in ANTICIPATE would therefore also have an anticipated additional costs of recommending $r_1 \in \mathcal{B}$ in R1. In consequence, only the exact, numerical prediction but none of the following qualitative predictions for O in ANTICIPATE change (formal results are available upon request).

The payoffs of all possible sequences of actions for advisers with $r^* \in \mathcal{B}$ are as follows:

$$\mathcal{BBB}: \quad u(b) + u(v(o^*)) - \lambda \cdot P_{\mathcal{BB}} - \lambda \cdot P_{\mathcal{BBB}} \tag{11}$$
$$\mathcal{BBN}: \quad u(b) + u(v(o^*)) - \lambda \cdot P_{\mathcal{BB}} - \kappa - \lambda \cdot P_{\mathcal{BBN}} \tag{12}$$
$$\mathcal{BNN}: \quad u(b) + u(v(o^*_{\mathcal{N}})) - \lambda \cdot P_{\mathcal{BN}} - \kappa - \lambda \cdot P_{\mathcal{BNN}} \tag{13}$$
$$\mathcal{BNB}: \quad u(b) + u(v(o^*_{\mathcal{N}})) - \lambda \cdot P_{\mathcal{BN}} - \lambda \cdot P_{\mathcal{BNB}} \tag{14}$$
$$\mathcal{NNN}: \quad u(v(o^*_{\mathcal{N}})) - 2\kappa \tag{15}$$
$$\mathcal{NNB}: \quad u(v(o^*_{\mathcal{N}})) - \kappa \tag{16}$$
$$\mathcal{NBN}: \quad u(v(o^*)) - 2\kappa \tag{17}$$
$$\mathcal{NBB}: \quad u(v(o^*)) - \kappa \tag{18}$$

For advisers with $r^* \in \mathcal{N}$ the following payoffs emerge:

$$\mathcal{BBB}: \quad u(b) - \kappa + u(v(o^*_{\mathcal{B}})) - \lambda \cdot P_{\mathcal{BB}} - \kappa - \lambda \cdot P_{\mathcal{BBB}} \tag{19}$$
$$\mathcal{BBN}: \quad u(b) - \kappa + u(v(o^*_{\mathcal{B}})) - \lambda \cdot P_{\mathcal{BB}} - \lambda \cdot P_{\mathcal{BBN}} \tag{20}$$
$$\mathcal{BNN}: \quad u(b) - \kappa + u(v(o^*)) - \lambda \cdot P_{\mathcal{BN}} - \lambda \cdot P_{\mathcal{BNN}} \tag{21}$$
$$\mathcal{BNB}: \quad u(b) - \kappa + u(v(o^*)) - \lambda \cdot P_{\mathcal{BN}} - \kappa - \lambda \cdot P_{\mathcal{BNB}} \tag{22}$$
$$\mathcal{NNN}: \quad u(v(o^*)) \tag{23}$$
$$\mathcal{NNB}: \quad u(v(o^*)) - \kappa \tag{24}$$
$$\mathcal{NBN}: \quad u(v(o^*_{\mathcal{B}})) \tag{25}$$
$$\mathcal{NBB}: \quad u(v(o^*_{\mathcal{B}})) - \kappa \tag{26}$$

Comparing these payoffs then leads to Proposition 4. It shows that the share of advisers with $r^* \in \mathcal{N}$ who choose $r_1 \in \mathcal{B}$ when they can anticipate upcoming actions is lower than the corresponding share when such anticipation is not possible. It also shows that in contrast to the setting where advisers could not anticipate upcoming actions, some advisers with $r^* \in \mathcal{B}$ do not plan to recommend $r_1 \in \mathcal{B}$ when they can anticipate upcoming choices. The intuition behind this result can be derived in four steps.

First, in a plan which features $r_1 \in \mathcal{B}$ and $o \in \mathcal{B}$, an adviser with $r^* \in \mathcal{B}$ would always prefer to recommend $r_2 \in \mathcal{B}$ rather than $r_2 \in \mathcal{N}$, because recommending the former avoids incurring the costs of giving inappropriate advice and he is less likely to be inferred as a type-3-adviser. Advisers with $r^* \in \mathcal{N}$ in the same context face a trade-off between the costs of giving inappropriate advice and image concerns for their plan regarding r_2. Such advisers either choose $\mathcal{BBB}$ or $\mathcal{BBN}$, depending on the relative size of their κ and λ. In any case, $P_{\mathcal{BBN}} = 1$ applies.

Second, given a plan which features $r_1 \in \mathcal{B}$ and $r_2 \in \mathcal{N}$, advisers with $r^* \in \mathcal{B}$ would always prefer $o \in \mathcal{B}$, because it entails both higher expected monetary payoff and less image concerns. This implies that these advisers, if their plan features $r_1 \in \mathcal{B}$, also plan to choose $o \in \mathcal{B}$ and that they prefer $\mathcal{BBB}$ over $\mathcal{BBN}$ and $\mathcal{BNN}$. Alternatively, advisers with $r^* \in \mathcal{N}$ in the same context face a trade-off between the expected monetary payoff and image concerns and hence, may choose either $\mathcal{BNN}$ or $\mathcal{BBN}$. Therefore, we have $P_{\mathcal{BN}} = 1$ since the above reasoning implies that advisers with $r^* \in \mathcal{B}$ never choose $o \in \mathcal{N}$, given that they plan to recommend $r_1 \in \mathcal{B}$ but $r_2 \in \mathcal{N}$.

Third, given that $P_{\mathcal{BN}} = 1$, advisers with $r^* \in \mathcal{B}$ prefer $\mathcal{BBB}$ over $\mathcal{BNB}$ because choosing the latter reveals such advisers to be of type-3. However, choosing the former induces a probability strictly less than one and higher expected monetary payoff. Advisers with $r^* \in \mathcal{N}$ would also never choose $\mathcal{BNB}$ as it is dominated by $\mathcal{BNN}$. This is because a choice $o \in \mathcal{N}$ following a recommendation $r_1 \in \mathcal{B}$ also reveals the adviser to be of type-3 while recommending $r_2 \in \mathcal{B}$ does not elevate image concerns. Taken together, this means that advisers with $r^* \in \mathcal{B}$ prefer $\mathcal{BBB}$ over the rest of plans which feature $r_1 \in \mathcal{B}$ while advisers with $r^* \in \mathcal{N}$ may prefer $\mathcal{BBB}$, $\mathcal{BBN}$, or $\mathcal{BNN}$ over other plans which feature $r_1 \in \mathcal{B}$.

Finally, after recommending $r_1 \in \mathcal{N}$, advisers with $r^* \in \mathcal{B}$ prefer $\mathcal{NBB}$ over other plans which feature $r_1 \in \mathcal{N}$. Alternatively, advisers with $r^* \in \mathcal{N}$ prefer $\mathcal{NNN}$ over the rest of plans which feature $r_1 \in \mathcal{N}$. These decisions are made because recommeding $r_1 \in \mathcal{N}$ eliminates the possibility that the adviser is type-3 and hence, all advisers would choose o^* in O and recommend r^* in R2.

Summarizing the above points, advisers with $r^* \in \mathcal{B}$ choose either $\mathcal{BBB}$ or $\mathcal{NBB}$ among all possible plans. They do not strictly prefer the former because this signals that this adviser is type-3 with strictly positive probability whereas the latter puts this posterior probability to zero. Alternatively, advisers with $r^* \in \mathcal{N}$ may choose one of the following four plans: $\mathcal{BBB}$, $\mathcal{BBN}$, $\mathcal{BNN}$, and $\mathcal{NNN}$. Note that the advisers who prefer the first three plans are, by definition, type-3, and the advisers who prefer the last plan are type-2. This leads to Proposition 4:

Proposition 4. *Suppose advisers consider choices in ANTICIPATE as multi-stage decision. Then, the share of advisers with $r^* \in \mathcal{B}$ who plan to choose $r_1 \in \mathcal{B}$ is given by $\tilde{\tau}^*_{r_1} \leq 1$. The share of advisers with $r^* \in \mathcal{N}$ who plan to choose $r_1 \in \mathcal{B}$ is given by $\tilde{\pi}^*_{r_1} < K(u(b))$.*

Proof: see Appendix B.

The first part of the above shows that the share of advisers with $r^* \in \mathcal{B}$ who plan to recommend $r_1 \in \mathcal{B}$ can be less than one (different to behavior when there is no possibility to anticipate and where this share equals one). In contrast, in the NO BONUS and BONUS treatments, all of these advisers – whose overall share in the population is given by $\beta_{\mathcal{B}}$ – make such a recommendation (see results for R1 in the preceding subsection). This is the effect of the anticipated image concerns for advisers with $r^* \in \mathcal{B}$: If such concerns are high enough, they want to avoid recommending $r_1 \in \mathcal{B}$ to rule out the possibility of being perceived as type-3s. The second part of the above proposition demonstrates that the share of advisers with $r^* \in \mathcal{N}$ who plan to recommend $r_1 \in \mathcal{B}$, just to earn the bonus, is decreased due to anticipated image costs. Whereas for unanticipated actions, their population share was given by $(1-\beta_{\mathcal{B}})K(u(b))$, the costs from such anticipated actions implies a smaller fraction of advisers planning to behave in this manner. Together, this yields the following key result which leads to Prediction 4b in the main text:

Corollary 4b. *If advisers consider choices in ANTICIPATE as multi-stage decisions, the share of them who recommend $r_1 \in \mathcal{B}$ in the ANTICIPATE treatment, given by $\beta_{\mathcal{B}} \cdot \tilde{\tau}^*_{r_1} + (1-\beta_{\mathcal{B}}) \cdot \tilde{\pi}^*_{r_1}$, is strictly lower than the share of them in BONUS, given by $\beta_{\mathcal{B}} + (1-\beta_{\mathcal{B}})K(u(b))$.*

A4. Competing explanations

The above analysis of behavior in our treatments illustrates how, through image concerns of being perceived as biased, a one-off bonus can lead advisers to repeated biased advice. It can even lead them to choose for themselves in a way which, absent such concerns, would be sub-optimal. One could argue that other explanations, e.g., anchoring-based explanation may also be consistent with some of findings.

The general idea of such theories is that the bonus in ANTICIPATE and BONUS is perceived as a signal or cue about which option is best or ought to be chosen and recommended. Thus, there is a shift of the mass of advisers who have r^* and o^* in $\mathcal{B}$ in these treatments. This means hat some advisers who, absent a bonus, do not have $r^* \in \mathcal{B}$ choose to recommend $r_1 \in \mathcal{B}$ not because of the pecuniary value of the bonus but because the cue it entails. Denote the share of such advisers by $\delta > 0$. Without image concerns, such advisers make every decision according to their (shifted) values for r^* and o^* are. Then, together with those who recommend $r_1 \in \mathcal{B}$ just for the bonus, i.e., type-3 advisers (with mass of ϕ_3), the difference in the total share of advisers who recommend $r_1 \in \mathcal{B}$ between BONUS and NO BONUS is given by $\delta + \phi_3$. The share of advisers δ whose preference is affected by the cue-effect of the bonus then also choose $o \in \mathcal{B}$, regardless of whether r^* and o^* are independent or not. Similarly, these advisers also recommend $r_2 \in \mathcal{B}$ in R2. In NO BONUS, when no such cue was there, they choose according to their original preference. Hence, column 2 of Table A.1 follows for the comparison of BONUS and NO BONUS. Comparing BONUS to ANTICIPATE, however, the anchoring based explanation predicts no differences. This is because both treatments feature the bonus and, therefore, also any cue or anchor it might entail. In consequence, column 5 in Table A.1 follows.

The predictions based on image concerns are also given in the table, in particular, column 1, 3, and 4. The predictions in column 1 are based on Corollary 1, 2, and 3. The predictions in column 3 are based on Corollary 4a. Finally, the predictions in column 4 are based on Corollary 4b. Recall that when advisers consider choices in ANTICIPATE as multi-stage decisions, the only prediction one can establish is that fewer advisers recommend $r_1 \in \mathcal{B}$ due to the anticipated image concerns (except for the "0" in the column's fourth row; see footnote 3).

In summary, both the anchoring based explanation and the image concerns theory predicts a treatment effect between BONUS and NO BONUS.[4] However, the different treatment effects between BONUS and ANTICIPATE clearly rejects the anchoring based explanation as it predicts no difference. The image concerns explanation, on the other hand, predicts a difference in either the share of choosing $o \in \mathcal{B}$ and recommending $r_2 \in \mathcal{B}$ (when choices are considered as one-stage decisions) or the share of recommending $r_1 \in \mathcal{B}$ (when choices are considered as multi-stage decisions).

	BONUS $-$ NO BONUS		BONUS $-$ ANTICIPATE		
	Image concerns	Anchoring & cue-effects	Image concerns		Anchoring & cue-effects
			One-stage	Multi-stage	
	(1)	(2)	(3)	(4)	(5)
R1	ϕ_3	$\delta + \phi_3$	0	$\phi_1(1 - \tilde{\tau}^*_{r_1}) + (1-\phi_1) \cdot [K(u(b)) - \tilde{\pi}^*_{r_1}]$	0
O if $r^* \perp\!\!\!\perp o^*$	0	δ	0	0	0
O if $r^* = o^*$	$\phi_3 \cdot \pi^*_o$	δ	$\phi_3 \cdot \pi^*_o$	n.a.	0
R2	$\phi_3 \cdot \pi^*_o \pi^*_{r_2}$	δ	$\phi_3 \cdot \pi^*_o \pi^*_{r_2}$	n.a.	0

Table A.1. Predicted differences in percentage of recommending/choosing from $\mathcal{B}$; n.a. denotes cases where no clear-cut prediction can be made (see footnote 17 in the main text)

[4]From comparing column 1 and 2 in the table, it is clear that the image concerns predict partial consistency (share π^*_o and $\pi^*_o \pi^*_{r_2}$ re-recommend and choose from $\mathcal{B}$) whereas the anchoring based explanation predicts full consistency for those who take the bonus as cue or anchor through it (share δ).

Appendix B: Formal results and proofs

Proof of Lemma 1

Let j be the joint p.d.f. associated with the joint c.d.f. J for (κ, λ). Accordingly, it holds that

$$K(x) = \int_0^x \int_0^\infty j(\kappa, \lambda) d\lambda d\kappa.$$

Full support for the joint p.d.f. j, i.e., $j(\kappa, \lambda) > 0$ for all $(\kappa, \lambda) \in \mathbb{R}_0^+ \times \mathbb{R}_0^+$, implies

$$K'(x) = \int_0^\infty j(x, \lambda) d\lambda > 0.$$

Repeating this for Λ proves part a). Part b) can be proven analogously, as for any $x > 0$

$$\begin{aligned} &\Lambda(y \mid x) = \Pr[\lambda \leq y \mid \kappa \leq x] = \left(\int_0^y \int_0^x j(\kappa, \lambda) d\kappa d\lambda\right) \Big/ \left(\int_0^\infty \int_0^x j(\kappa, \lambda) d\kappa d\lambda\right) \\ \Rightarrow \quad &\Lambda'(y \mid x) = \left(\int_0^x j(\kappa, y) d\kappa\right) \Big/ \left(\int_0^\infty \int_0^x j(\kappa, \lambda) d\kappa d\lambda\right) > 0. \end{aligned}$$

For part c), rewrite the conditions $\kappa \in [0, x]$ and $\kappa/\lambda \leq z$ as $\kappa \in \{0, \min\{x, z\lambda\}\}$. We can then write the conditional c.d.f. $R(z \mid x, y) = \Pr[\kappa/\lambda \leq z \mid \kappa \leq x, \lambda \geq y]$ with $x > 0$ as

$$R(z \mid x, y) = \left(\int_y^\infty \int_0^{\min\{x, z\lambda\}} j(\kappa, \lambda) d\kappa d\lambda\right) \Big/ \left(\int_y^\infty \int_0^x j(\kappa, \lambda) d\kappa d\lambda\right)$$

If $x > z\lambda$, the partial derivative of the above w.r.t. z is then given by

$$R'(z \mid x, y) = \left(\int_y^\infty \lambda \cdot j(z\lambda, \lambda) d\lambda\right) \Big/ \left(\int_y^\infty \int_0^x j(\kappa, \lambda) d\kappa d\lambda\right)$$

and strictly positive by full support of j. If $x \leq z\lambda$, the above equals zero so that $R'(z \mid x, y) \geq 0$ holds. □

Proof of Proposition 1

First note that for type-2-advisers, $r_1 \in \mathcal{N}$. Since type-3-advisers have $r_1 \in \mathcal{B}$, $\Pr[\theta = 3 \mid r_1 \in \mathcal{N}, o] = 0$; type-2-advisers cannot be perceived as type-3s. Type-2-advisers therefore maximize $U(o \mid r_1 \in \mathcal{N}, r^*) = u(v(o))$ by choosing $o^* = r^* = r_1 \in \mathcal{N}$ for themselves. In contrast, advisers of type $\theta \in \{1, 3\}$ have recommended $r_1 \in \mathcal{B}$ such that they both can be inferred to be possibly of type $\theta = 3$. Suppose share τ_o of type-1-advisers choose for themselves such that $o \in \mathcal{B}$. Similarly, let π_o denote the share of type-3-advisers who choose for themselves $o \in \mathcal{B}$. The following posteriors then emerge:

$$\Pr[\theta = 3 \mid o \in \mathcal{B}, r_1 \in \mathcal{B}] = \frac{\pi_o \cdot \phi_3}{\tau_o \cdot \phi_1 + \pi_o \cdot \phi_3} \tag{27}$$

$$\Pr[\theta = 3 \mid o \in \mathcal{N}, r_1 \in \mathcal{B}] = \frac{(1 - \pi_o) \cdot \phi_3}{(1 - \tau_o) \cdot \phi_1 + (1 - \pi_o) \cdot \phi_3} \tag{28}$$

It is easily verified that the latter posterior is weakly larger than the former if and only if $\tau_o \geq \pi_o$. If this condition applies, then it holds for type-1-advisers (for whom $o^* = o^*_{\mathcal{B}}$) that for any $o' \in \mathcal{N}$

$$u(v(o')) - \lambda \cdot \Pr[\theta = 3 \mid o', r_1 \in \mathcal{B}] < u(v(o^*)) - \lambda \cdot \Pr[\theta = 3 \mid o^*, r_1 \in \mathcal{B}].$$

If type-1-advisers chose $o' \in \mathcal{N}$ they would suffer for two reasons: First, such choices are suboptimal in terms of maximizing their pecuniary utility $u(v(o))$. Second, choosing $o' \in \mathcal{N}$ leads to a worse image utility through a higher probability to be perceived as type-3. Accordingly, in all equilibria with $\tau_o \geq \pi_o$ all type-1-advisers choose $o = o^* \in \mathcal{B}$. Therefore, $\tau_o = 1 \geq \pi_o$ has to hold for all equilibria in this class.

In the candidate equilibrium with $\tau_o = 1 \geq \pi_o$, all type-1-advisers choose $o = o^* = r^* = r_1 \in \mathcal{B}$. Type-3-advisers can thus pool with type-1s by choosing consistently from $\mathcal{B}$, i.e., $o = r_1 \in \mathcal{B}$, even though for them $o^* \in \mathcal{N}$. They then choose their constrained optimum $o^*_{\mathcal{B}} \in \mathcal{B}$. If they do not choose consistently they can choose their preferred option $o^* \in \mathcal{N}$ but will then reveal themselves as biased, i.e., as type-3-advisers. Using (27) and the assumption that in case of indifference they choose o^*, this means that type-3-advisers pool if the following holds:

$$u(v(o^*)) - \lambda < u(v(o^*_{\mathcal{B}})) - \lambda \cdot \frac{\pi_o \phi_3}{\phi_1 + \pi_o \phi_3} \quad \Leftrightarrow \quad \lambda > (u(v(o^*)) - u(v(o^*_{\mathcal{B}}))) \cdot \left(\frac{\phi_1 + \pi_o \phi_3}{\phi_1} \right)$$

Since for type-3-advisers $u(v(o^*)) > u(v(o^*_{\mathcal{B}}))$, the threshold on the RHS of the second inequality grows in π_o, the share of type-3-advisers who choose $o \in \mathcal{B}$ to pool with type-1s. In addition, because they are type-3-advisers, $\kappa < u(b)$ has to hold. Therefore, the share π_o of pooling type-3-advisers has to solve

$$1 - \pi_o = \Lambda \left((u(v(o^*)) - u(v(o^*_{\mathcal{B}}))) \cdot \left(\frac{\phi_1 + \pi_o \phi_3}{\phi_1} \right) \Big| \, u(b) \right).$$

From Lemma 1 b), it follows immediately that both $\pi_o = 0$ and $\pi_o = 1$ cannot be solutions. Also, the above RHS is strictly increasing in π_o while its values are contained in the unit interval. The above LHS is simply the decreasing 45-degree-line over the unit square. Accordingly, there has to be a unique solution $\pi^*_o \in (0, 1)$.

Finally, we exclude other equilibria with $\tau_o < \pi_o$. In this case, the posterior (27) is strictly larger than (28). Since for type-3-advisers $o^* = o^*_{\mathcal{N}}$ holds, they then choose their preferred choice as

$$u(v(o^*)) - \lambda \cdot \Pr[\theta = 3 \mid o^*, r_1 \in \mathcal{B}] > u(v(o^*_{\mathcal{B}})) - \lambda \cdot \Pr[\theta = 3 \mid o^*_{\mathcal{B}}, r_1 \in \mathcal{B}].$$

Thus, all type-3-advisers choose $o \in \mathcal{N}$ and reveal themselves. This corresponds to $\pi_o = 0$ and therefore contradicts an equilibrium with $\tau_o < \pi_o$. □

Proof of Proposition 2

Type-2-advisers have initially recommended $r_1 \in \mathcal{N}$. As type-3-advisers have recommended $r_1 \in \mathcal{B}$, it holds that $\Pr[\theta = 3 \mid r_1 \in \mathcal{N}, o, r_2] = 0$ and type-2s therefore maximize $U(o \mid r_1 \in \mathcal{N}, r^*) = -\kappa \cdot \mathbb{1}[r_2 \neq r^*]$ by re-recommending $r_2 = r^* = r_1 \in \mathcal{N}$.

We now look on type-3-advisers. First, consider the situation that appropriate recommendations are projected from own choice ($r^* = o^*$). Share $1 - \pi^*_o \in (0, 1)$ of type-3-advisers has then already revealed themselves as such by choosing $o \neq r_1$ in the own choice O (see Lemma 1). Therefore, their image concerns are invariant to r_2 as for them, $\Pr[\theta = 3 \mid o \neq r_1, r_2] = 1$ applies for every $r_2 \in \mathcal{C}$. They then maximize $U(r_2 \mid r_1, r_2, r^*) = -\kappa \cdot \mathbb{1}[r_2 \neq r^*] - \lambda \cdot \Pr[\theta = 3 \mid r_2, r_1, o]$ by recommending $r_2 = r^* \in \mathcal{N}$. Therefore, they do not re-recommend their initial, biased recommendation $r_1 \in \mathcal{B}$.

Type-1-advisers and share $\pi_o \in (0, 1)$ of type-3-advisers who have not yet revealed themselves both look identical to an outside observer as both have a history of $o = r_1 \in \mathcal{B}$. Accordingly, hitherto unrevealed type-3-advisers can continue to pool with type-1-advisers. Denote with τ_{r_2} the share of type-1-advisers who recommend $r_2 \in \mathcal{B}$ and with π_{r_2} the share of type-3-advisers who recommend $r_2 \in \mathcal{B}$. This yields the following posteriors, conditional on not having previously revealed oneself (i.e., that $o = r_1$ holds):

$$\Pr[\theta = 3 \mid r_2 \in \mathcal{B}, r_1 \in \mathcal{B}] = \frac{\pi_{r_2} \cdot \pi^*_o \phi_3}{\tau_{r_2} \phi_1 + \pi_{r_2} \cdot \pi^*_o \phi_3} \tag{29}$$

$$\Pr[\theta = 3 \mid r_2 \in \mathcal{N}, r_1 \in \mathcal{B}] = \frac{(1 - \pi_{r_2}) \cdot \pi^*_o \phi_3}{(1 - \tau_{r_2}) \cdot \phi_1 + (1 - \pi_{r_2}) \cdot \pi^*_o \phi_3} \tag{30}$$

Posterior (30) is weakly larger than (29) if and only if $\tau_{r_2} \geq \pi_{r_2}$. If this condition holds, the payoff for type-1-advisers with $r^* \in \mathcal{B}$ from re-recommending r^* in R2 is always strictly larger than from recommending $r_2' \in \mathcal{N}$:

$$-\lambda \cdot \Pr[\theta = 3 \mid r^*, r_1 \in \mathcal{B}] > -\kappa - \lambda \cdot \Pr[\theta = 3 \mid r_2', r_1 \in \mathcal{B}]$$

Thus, the only equilibrium with $\tau_{r_2} \geq \pi_{r_2}$ obeys $\tau_{r_2} = 1$. Hitherto unrevealed type-3-advisers who want to pool with type-1s have to choose analogously, i.e., $r_2 = r_1 \in \mathcal{B}$, even though this is not their appropriate choice because for them, $r^* \in \mathcal{N}$ holds. This allows them to not reveal themselves as biased so that their image costs λ are discounted by $\Pr[\theta = 3 \mid r_2 \in \mathcal{B}, o = r_1 \in \mathcal{B}]$. For this, they experience costs κ of recommending something they do not consider appropriate. Plugging in the above posteriors with $\tau_{r_2} = 1$ yields

$$-\lambda < -\kappa - \lambda \cdot \frac{\pi_{r_2} \cdot \pi_o^* \phi_3}{\phi_1 + \pi_{r_2} \cdot \pi_o^* \phi_3} \quad \Leftrightarrow \quad \kappa < \lambda \cdot \frac{\phi_1}{\phi_1 + \pi_{r_2} \cdot \pi_o^* \phi_3}$$

as a condition for hitherto unrevealed type-3-advisers to continue pooling with type-1s. Note that Lemma 1b implies that for every κ multiplied with some factor, there is a mass of advisers with sufficiently high λ, i.e., with $\lambda > \kappa(\phi_1 + \pi_{r_2} \cdot \pi_o^* \phi_3)/\phi_1$. Therefore, $\pi_{r_2} = 0$ cannot be true. Also, the limit on κ/λ which the above inequality implies is only relevant to type-3-advisers (those advisers who have $\kappa < u(b)$) who have not revealed themselves in O (those with $\lambda > (u(v(o^*)) - u(v(o^*_{\mathcal{B}}))) \cdot (\phi_1 + \pi_o^* \phi_3)/\phi_1$). Thus, the share of type-3-advisers who continue to pool, denoted by π_{r_2}, is determined by the solution to

$$\pi_{r_2} = R\left(\frac{\phi_1}{\phi_1 + \pi_{r_2} \cdot \pi_o^* \phi_3} \;\middle|\; u(b), (u(v(o^*)) - u(v(o^*_{\mathcal{B}}))) \cdot \left(\frac{\phi_1 + \pi_o^* \phi_3}{\phi_1}\right)\right). \tag{31}$$

A solution $\pi_{r_2} = 0$ has been ruled out above. By Lemma 1c the above RHS is non-increasing in π_{r_2} and takes a value in the unit interval. As the LHS is just the 45-degree line above it, there has to be a unique intersection for some $\pi^*_{r_2} \in (0, 1]$.

We exclude equilibria with $\tau_{r_2} < \pi_{r_2}$ in a similar fashion as in the proof of Lemma 1. With $\tau_{r_2} < \pi_{r_2}$, the posterior (29) is larger than (30). For hitherto unrevealed type-3-advisers with $r^* \in \mathcal{N}$ it thus holds that

$$-\kappa - \lambda \cdot \Pr[\theta = 3 \mid r_2 \in \mathcal{B}, r_1 \in \mathcal{B}] < -\lambda \cdot \Pr[\theta = 3 \mid r^* \in \mathcal{N}, r_1 \in \mathcal{B}]$$

and they all recommend the choice $r^* \in \mathcal{N}$ and thereby reveal themselves. This implies $\pi_{r_2} = 0$ and thus contradicts an equilibrium with $\tau_{r_2} < \pi_{r_2}$.

Recall from Lemma 1 and its proof that when what is cosidered appropriate advice is independent from own choices, o has no diagnostic value and type-3-advisers have not had yet the possibility to reveal themselves in O. In terms of signaling value for R2, this is equivalent to the above when $\pi_o^* = 1$. The above reasoning can then be repeated with this parameter choice when the RHS in (31) is replaced by $R\left(\phi_1/(\phi_1 + \pi_{r_2}\phi_3) \mid u(b), 0\right)$ as no prior chance to reveal oneself does not restrict the subset of those type-3-advisers who can pool (i.e., it does not restrict the values of λ). The qualitative results, however, remain unchanged. □

Proof of Proposition 4

We prove this proposition through a series of lemmas which sequentially rule out plans of actions for different adviser types.

Lemma 2. *Advisers with $r^* \in \mathcal{B}$ prefer $\mathcal{BBB}$ over $\mathcal{BBN}$ (i.e., (11) over (12)) while there is a positive share of advisers with $r^* \in \mathcal{N}$ who prefer $\mathcal{BBB}$ over $\mathcal{BBN}$ (i.e., (19) over (20)) and there is a positive share of such advisers who prefer $\mathcal{BBN}$ over $\mathcal{BBB}$ (i.e., (20) over (19)). Furthermore, $P_{\mathcal{BBN}} = 1$.*

Proof. The relevant posteriors are given below, where π_o, π_{r_2}, τ_o, and τ_{r_2} are defined as in the proof of Proposition 1 and 2:

$$P_{\mathcal{BBB}} = \Pr[\theta = 3 \mid r_2 \in \mathcal{B}, o \in \mathcal{B}, r_1 \in \mathcal{B}] \quad = \quad \frac{\pi_{r_2} \cdot \pi_o \phi_3}{\tau_{r_2} \cdot \tau_o \phi_1 + \pi_{r_2} \cdot \pi_o \phi_3} \tag{32}$$

$$P_{\mathcal{BBN}} = \Pr[\theta = 3 \mid r_2 \in \mathcal{N}, o \in \mathcal{B}, r_1 \in \mathcal{B}] \quad = \quad \frac{(1 - \pi_{r_2}) \cdot \pi_o \phi_3}{(1 - \tau_{r_2}) \cdot \tau_o \phi_1 + (1 - \pi_{r_2}) \cdot \pi_o \phi_3} \tag{33}$$

Posterior (33) is weakly larger than (32) if and only if $\tau_{r_2} \geq \pi_{r_2}$. If this condition holds, the payoff for advisers with $r^* \in \mathcal{B}$ from re-recommending r^* in R2 is always strictly larger than from recommending $r_2 \in \mathcal{N}$, i.e.,

$$-\lambda \cdot P_{\mathcal{BBB}} > -\kappa - \lambda \cdot P_{\mathcal{BBN}}.$$

Thus, as long as advisers with $r^* \in \mathcal{B}$ plan to choose $o \in \mathcal{B}$, the only equilibrium with $\tau_{r_2} \geq \pi_{r_2}$ obeys $\tau_{r_2} = 1$. In this case the payoff of choosing $\mathcal{BBB}$ is strictly larger than the payoff of choosing $\mathcal{BBN}$.

To exclude equilibria with $\tau_{r_2} < \pi_{r_2}$, suppose to the contrary that this condition held. Then the posterior (33) is smaller than (32). Those advisers with $r^* \in \mathcal{N}$ who prefer $\mathcal{BBN}$ over $\mathcal{BBB}$ must then have

$$-\kappa - \lambda \cdot P_{\mathcal{BBB}} < -\lambda \cdot P_{\mathcal{BBN}}$$

which means that they all recommend from $\mathcal{N}$ and thereby reveal themselves. This implies $\pi_{r_2} = 0$ and thus contradicts an equilibrium with $\tau_{r_2} < \pi_{r_2}$.

Plugging $\tau_{r_2} = 1$ into the posteriors we get $P_{\mathcal{BBB}} = \frac{\pi_{r_2} \cdot \pi_o \phi_3}{\tau_o \phi_1 + \pi_{r_2} \cdot \pi_o \phi_3}$ and $P_{\mathcal{BBN}} = 1$. Therefore, advisers with $r^* \in \mathcal{N}$ choose $\mathcal{BBB}$ if and only if

$$-\kappa - \lambda \cdot P_{\mathcal{BBB}} > -\lambda \cdot P_{\mathcal{BBN}} \quad \Leftrightarrow \quad \kappa < \lambda \cdot \frac{\tau_o \phi_1}{\tau_o \phi_1 + \pi_{r_2} \cdot \pi_o \phi_3}. \tag{34}$$

For advisers with $r^* \in \mathcal{N}$ who plan to recommend $r_1 \in \mathcal{B}$ and to choose $o \in \mathcal{B}$ it then follows from Lemma 1c that there is a positive mass of them who choose $\mathcal{BBB}$ and a positive mass who choose $\mathcal{BBN}$ (i.e., those for whom the above inequality does hold or does not hold, respectively). □

While the preceding lemma refers to the second recommendation, Lemmas 3 and 4 pin down advisers' own choices:

Lemma 3. *Advisers with $r^* \in \mathcal{B}$ prefer $\mathcal{BBN}$ over $\mathcal{BNN}$ (i.e, (4) over (6)), while there is a positive share of advisers with $r^* \in \mathcal{N}$ who prefer $\mathcal{BBN}$ over $\mathcal{BNN}$ (i.e.,(21) over (20)) and a positive share of such advisers who prefer $\mathcal{BNN}$ over $\mathcal{BBN}$ (i.e.,(20) over (21)). Furthermore, $P_{\mathcal{BNN}} = P_{\mathcal{BN}} = 1$.*

Proof. The relevant posteriors are given by

$$P_{\mathcal{BB}} = \Pr[\theta = 3 \mid o \in \mathcal{B}, r_1 \in \mathcal{B}] \quad = \quad \frac{\pi_o \cdot \phi_3}{\tau_o \cdot \phi_1 + \pi_o \cdot \phi_3}$$

$$P_{\mathcal{BN}} = \Pr[\theta = 3 \mid o \in \mathcal{N}, r_1 \in \mathcal{B}] \quad = \quad \frac{(1 - \pi_o) \cdot \phi_3}{(1 - \tau_o) \cdot \phi_1 + (1 - \pi_o) \cdot \phi_3}$$

Again, it is easily verified that the latter is weakly larger than the former if and only if $\tau_o \geq \pi_o$. If this condition applies, then it must be true that advisers with $r^* \in \mathcal{B}$ prefer $\mathcal{BBN}$ over $\mathcal{BNN}$ since

$$u(o^*) - \lambda \cdot P_{\mathcal{BB}} - \lambda \cdot P_{\mathcal{BBN}} > u(o^*_{\mathcal{B}}) - \lambda \cdot P_{\mathcal{BN}} - \lambda \cdot P_{\mathcal{BNN}}. \tag{35}$$

The above follows from the fact that $P_{\mathcal{BBN}} = P_{\mathcal{BNN}} = 1$. To see this, consider these two posteriors:

$$\begin{aligned} P_{\mathcal{BBN}} = \Pr[\theta = 3 \mid r_2 \in \mathcal{N}, o \in \mathcal{B}, r_1 \in \mathcal{B}] &= \frac{(1-\pi_{r_2}) \cdot \pi_o \phi_3}{(1-\tau_{r_2}) \cdot \tau_o \phi_1 + (1-\pi_{r_2}) \cdot \pi_o \phi_3} \\ P_{\mathcal{BNN}} = \Pr[\theta = 3 \mid r_2 \in \mathcal{N}, o \in \mathcal{N}, r_1 \in \mathcal{B}] &= \frac{(1-\pi_{r_2}) \cdot (1-\pi_o)\phi_3}{(1-\tau_{r_2}) \cdot (1-\tau_o)\phi_1 + (1-\pi_{r_2}) \cdot (1-\pi_o)\phi_3} \end{aligned}$$

It can be easily verified that $P_{\mathcal{BBN}} \leq P_{\mathcal{BNN}}$ if $\tau_o \geq \pi_o$. From Lemma 2, we get that $P_{\mathcal{BNN}} = 1$ which then implies $P_{\mathcal{BBN}} = 1$. As (35) holds, when then get $\tau_o = 1 \geq \pi_o$ for all equilibria with $\tau_o \geq \pi_o$. This also means $P_{\mathcal{BN}} = 1$.

Now we exclude equilibria with $\tau_o < \pi_o$. Suppose this were the case, then advisers with $r^* \in \mathcal{N}$ would strictly prefer $\mathcal{BNN}$ over $\mathcal{BBN}$ since

$$u(o^*_{\mathcal{B}}) - \lambda \cdot P_{\mathcal{BB}} < u(o^*) - \lambda \cdot P_{\mathcal{BN}}$$

where this comparison again uses the previous findings that $P_{\mathcal{BNN}} = P_{\mathcal{BBN}} = 1$. But this then implies that $\pi_o = 0$, a contradiction. Therefore, advisers with $r^* \in \mathcal{N}$ prefer $\mathcal{BBN}$ over $\mathcal{BNN}$ if and only if

$$u(o^*_{\mathcal{B}}) - \lambda \cdot P_{\mathcal{BB}} > u(o^*) - \lambda \;\; \Leftrightarrow \;\; \lambda > \frac{u(o^*) - u(o^*_{\mathcal{B}})}{1 - P_{\mathcal{BB}}}. \tag{36}$$

Given that $u(o^*) > u(o^*_{\mathcal{B}})$ and $P_{\mathcal{BB}} < 1$, we know that the RHS of the above inequality is strictly positive. According to full support assumption on the distribution of κ and λ, the probability that λ satisfies the above inequality is strictly positive and less than one. This implies that the mass of advisers with $r^* \in \mathcal{N}$ who prefer $\mathcal{BNN}$ over $\mathcal{BBN}$ is strictly positive and less than one, and that the mass of advisers with $r^* \in \mathcal{N}$ who prefer $\mathcal{BBN}$ over $\mathcal{BNN}$ is also strictly positive and less than one. □

Building on Lemma 3, the next lemma rules out two of the remaining options for advisers with $r^* \in \mathcal{B}$ and shows that they their plan never features $o \in \mathcal{N}$:

Lemma 4. *In every equilibrium, advisers with $r^* \in \mathcal{B}$ prefer $\mathcal{BBB}$ over $\mathcal{BNN}$ and $\mathcal{BNB}$ (i.e, (12) over (13) and (14)). Furthermore, $P_{\mathcal{BNB}} = 1$.*

Proof. Using $P_{\mathcal{BBN}} = P_{\mathcal{BNN}} = 1$, one gets from (13) that advisers with $r^* \in \mathcal{B}$ who choose $\mathcal{BNN}$ have a payoff of $u(b) + u(o^*_{\mathcal{B}}) - \lambda - \kappa - \lambda,$. This is lower than (11), the payoff from choosing $\mathcal{BBB}$ which is given by $u(b) + u(o^*) - \lambda \cdot P_{\mathcal{BB}} - \lambda \cdot P_{\mathcal{BBB}}$. In a similar manner, (14), the payoff from choosing $\mathcal{BNB}$, becomes $u(b) + u(o^*_{\mathcal{B}}) - \lambda - \lambda$ which is also strictly lower than (11). Therefore, all advisers with $r^* \in \mathcal{B}$ who plan to recommended $r_1 \in \mathcal{B}$ also choose $o \in \mathcal{B}$ with probability one, i.e., $\tau_o = 1$. In addition, we know that the following posterior must equal to one after plugging $\tau_o = 1$ in:

$$P_{\mathcal{BNB}} = \frac{\pi_{r_2} \cdot (1-\pi_o)\phi_3}{\tau_{r_2}(1-\tau_o)\phi_1 + \pi_{r_2} \cdot (1-\pi_o)\phi_3} = 1$$

□

Lemma 5. *In every equilibrium, advisers with $r^* \in \mathcal{N}$ prefer $\mathcal{BNN}$ over $\mathcal{BNB}$ (i.e., (21) over (22)).*

Proof. From Lemma 4, we know that $P_{\mathcal{BNB}} = 1$. This means that advisers with $r^* \in \mathcal{B}$ prefer $\mathcal{BNB}$ over $\mathcal{BNN}$. Advisers with $r^* \in \mathcal{N}$ prefer $\mathcal{BNN}$ over $\mathcal{BNB}$ if and only if

$$-\lambda \cdot P_{\mathcal{BNN}} \geq -\kappa - \lambda \cdot P_{\mathcal{BNB}}$$

which is always true given that $P_{\mathcal{BNB}} = P_{\mathcal{BNN}} = 1$. □

Finally, we provide the last lemma which regards plans featuring $r_1 \in \mathcal{B}$. It facilitates a comparison between $\mathcal{BNN}$ and $\mathcal{BBB}$.

Lemma 6. *In every equilibrium, a strictly positive mass of advisers with $r^* \in \mathcal{N}$ prefer $\mathcal{BBB}$ over $\mathcal{BNN}$, whereas a strictly positive mass of these advisers prefer $\mathcal{BNN}$ over $\mathcal{BBB}$.*

Proof. Advisers with $r^* \in \mathcal{N}$ prefer $\mathcal{BNN}$ over $\mathcal{BBB}$ if and only if (21) is no less than (19), i.e.,

$$\begin{aligned} & u(o^*) - 2\lambda \geq u(o^*_{\mathcal{B}}) - \lambda \cdot P_{\mathcal{BB}} - \kappa - \lambda \cdot P_{\mathcal{BBB}} \\ \Leftrightarrow \quad & \kappa \geq \lambda \cdot [1 - P_{\mathcal{BB}} + 1 - P_{\mathcal{BBB}}] - [u(o^*) - u(o^*_{\mathcal{B}})]. \end{aligned} \tag{37}$$

For λ small enough, the RHS of the second inequality given above must be no larger than zero, hence the inequality is satisfied for any $\kappa > 0$. This means there is a positive probability mass of (κ, λ) satisfies the inequality. This completes the proof. □

Now we determine which plans that feature $r_1 \in \mathcal{N}$ are preferred by advisers:

Lemma 7. *Advisers with $r^* \in \mathcal{B}$ who plan to choose $r_1 \in \mathcal{N}$ prefer $\mathcal{NBB}$ over $\mathcal{NNB}$, $\mathcal{NBN}$, and $\mathcal{NNN}$ (i.e., (18) over (16), (17), and (15)). Advisers with $r^* \in \mathcal{N}$ who plan to choose $r_1 \in \mathcal{N}$ prefer $\mathcal{NNN}$ over $\mathcal{NNB}$, $\mathcal{NBN}$, and $\mathcal{NBB}$ (i.e., (23) over (24), (25), and (26)).*

Proof. Lemma 7 follows from two comparisons: First, comparing (18) to (15), (16), and (17), respectively and second, comparing (23) to (24), (25), and (26), respectively. □

The results up to now show that advisers with $r^* \in \mathcal{B}$ prefer $\mathcal{BBB}$ in case they plan to initially recommended $r_1 \in \mathcal{B}$ and that they prefer $\mathcal{NBB}$ in case they plan to initially recommend $r_1 \in \mathcal{N}$. The percentage of advisers with $r^* \in \mathcal{B}$ who would recommend $r_1 \in \mathcal{B}$ hence depends on the percentage of them who prefer $\mathcal{BBB}$ over $\mathcal{NBB}$, i.e., for whom

$$u(b) + u(o^*) - \lambda \cdot P_{\mathcal{BB}} - \lambda \cdot P_{\mathcal{BBB}} \geq u(o^*) - \kappa \quad \Leftrightarrow \quad \kappa \geq \lambda \cdot (P_{\mathcal{BB}} + P_{\mathcal{BBB}}) - u(b).$$

holds. Therefore, the percentage of advisers with $r^* \in \mathcal{B}$ who choose $r_1 \in \mathcal{B}$ is given by

$$\tilde{\tau}^*_{r_1} \equiv \int_0^\infty \int_{\max\{0, \lambda \cdot (P_{\mathcal{BB}} + P_{\mathcal{BBB}}) - u(b)\}}^\infty j(\kappa, \lambda) d\kappa d\lambda \leq \int_0^\infty \int_0^\infty j(\kappa, \lambda) d\kappa d\lambda = 1.$$

The above proves the first part of the proposition. Advisers with $r^* \in \mathcal{N}$, on the other hand, may prefer $\mathcal{BBB}$, $\mathcal{BBN}$, or $\mathcal{BNN}$ if they plan to recommend $r_1 \in \mathcal{B}$. In case they plan to recommend $r_1 \in \mathcal{N}$, the previous results show that they only do so in the sequence $\mathcal{NNN}$. For advisers with $r^* \in \mathcal{N}$, the share of them who recommend $r_1 \in \mathcal{B}$ is therefore determined by comparing each of the plans $\mathcal{BBB}$, $\mathcal{BBN}$, and $\mathcal{BNN}$, to $\mathcal{NNN}$ if such a plan is most preferred among the plans which feature $r_1 \in \mathcal{B}$. For this, it is convenient to denote with Π_1, Π_2, and Π_3 the three partitions which divide the mass of advisers with $r^* \in \mathcal{N}$ and who recommend $r_1 \in \mathcal{B}$ in equilibrium:

Π_1: Advisers who prefer $\mathcal{BBB}$ over the rest of plans which feature $r_1 \in \mathcal{B}$ and over $\mathcal{NNN}$.

Π_2: Advisers who prefer $\mathcal{BBN}$ over the rest of plans which feature $r_1 \in \mathcal{B}$ and over $\mathcal{NNN}$.

Π_3: Advisers who prefer $\mathcal{BNN}$ over the rest of plans which feature $r_1 \in \mathcal{B}$ and over $\mathcal{NNN}$.

For advisers in Π_1, (34) and (36) must be satisfied for $\mathcal{BBB}$ being preferred over the rest of plans which feature $r_1 \in \mathcal{B}$. In addition, the following condition must hold to ensure that $\mathcal{BBB}$ is preferred over $\mathcal{NNN}$, i.e., that (19) is larger than (23):

$$\kappa < \frac{1}{2} \left[u(b) - [u(o^*) - u(o^*_{\mathcal{B}})] - \lambda \cdot (P_{\mathcal{BB}} + P_{\mathcal{BBB}}) \right]. \tag{38}$$

In consequence, Π_1 can also be defined via the restriction put on the (κ, λ)-values of the advisers:

$$\Pi_1 \equiv \{(\kappa, \lambda) \text{ s.t. conditions (34), (36), and (38) are satisfied}\}$$

Similarly, for advisers in Π_2, (36) and the opposite of (34) must be satisfied to ensure that $\mathcal{BBN}$ is preferred over the rest of plans which feature $r_1 \in \mathcal{B}$. In addition, the following condition must hold to ensure that $\mathcal{BBN}$ is preferred over $\mathcal{NNN}$, i.e., that (20) is larger than (23):

$$\kappa < u(b) - [u(o^*) - u(o^*_{\mathcal{B}})] - \lambda \cdot (P_{\mathcal{BB}} + 1). \tag{39}$$

This then allows to (re-)define Π_2 as follows:

$$\Pi_2 \equiv \{(\kappa, \lambda) \text{ s.t. condition (36), condition (39), and the opposite of condition (34) are satisfied}\}$$

For advisers in Π_3, (37) and the opposite of (34) must be satisfied to ensure that $\mathcal{BNN}$ is preferred over the rest of plans which feature $r_1 \in \mathcal{B}$. In addition, the following condition must hold to ensure that $\mathcal{BNN}$ is preferred over $\mathcal{NNN}$, i.e., that (21) is larger than (23):

$$\kappa < u(b) - 2 \cdot \lambda. \tag{40}$$

The share of these advisers in the population of advisers with $r^* \in \mathcal{N}$ is thus given by

$$\Pi_3 \equiv \{(\kappa, \lambda) \text{ s.t. condition (37), condition (40), and the opposite of condition (36) are satisfied}\}$$

.

For ease of exposition, denote by $k_{34}(\lambda)$ the RHS of (34), $k_{37}(\lambda)$ the RHS of (37), $k_{38}(\lambda)$ the RHS of (38), $k_{39}(\lambda)$ the RHS of (39), $k_{40}(\lambda)$ the RHS of (40), and l_{36} the RHS of (36). We can then compute the share of advisers within these three partitions as the share of the total population of advisers as follows:

$$\begin{aligned}
\tilde{\pi}^*_{r_1} &= \sum\nolimits_{t=1}^{3} \iint_{\Pi_t} j(\kappa,\lambda) d\kappa d\lambda \\
&= \int_{l_{36}}^{\infty} \int_0^{\min\{k_{34}(\lambda), k_{38}(\lambda)\}} j(\kappa,\lambda) d\kappa d\lambda + \int_{l_{36}}^{\infty} \int_{k_{34}(\lambda)}^{k_{39}(\lambda)} j(\kappa,\lambda) d\kappa d\lambda + \int_0^{l_{36}} \int_{\max\{k_{34}(\lambda), k_{37}(\lambda)\}}^{k_{40}(\lambda)} j(\kappa,\lambda) d\kappa d\lambda \\
&\leq \int_{l_{36}}^{\infty} \int_0^{k_{39}(\lambda)} j(\kappa,\lambda) d\kappa d\lambda + \int_0^{l_{36}} \int_{\max\{k_{34}(\lambda), k_{37}(\lambda)\}}^{k_{40}(\lambda)} j(\kappa,\lambda) d\kappa d\lambda \\
&< \int_{l_{36}}^{\infty} \int_0^{u(b)} j(\kappa,\lambda) d\kappa d\lambda + \int_0^{l_{36}} \int_0^{u(b)} j(\kappa,\lambda) d\kappa d\lambda \\
&= \int_0^{\infty} \int_0^{u(b)} j(\kappa,\lambda) d\kappa d\lambda = K(u(b)).
\end{aligned}$$

The inequalities are because the conditions (38), (39), and (40) restrict κ – and therefore, the respective upper limits on it in the above integrals – to be strictly less than $u(b)$. Also, $l_{36} > 0$, $k_{34}(\lambda) \geq 0$ holds for any $\lambda \geq 0$. Hence, the mass of advisers with $r^* \in \mathcal{N}$ who recommend $r_1 \in \mathcal{B}$ is lower than $K(u(b))$ which proves the second part of the proposition. □

Proof of Proposition 3

We use the following two lemmas which assume that advisers do not factor in image costs and are straightforward to prove from the payoffs stated in (3) through (10).

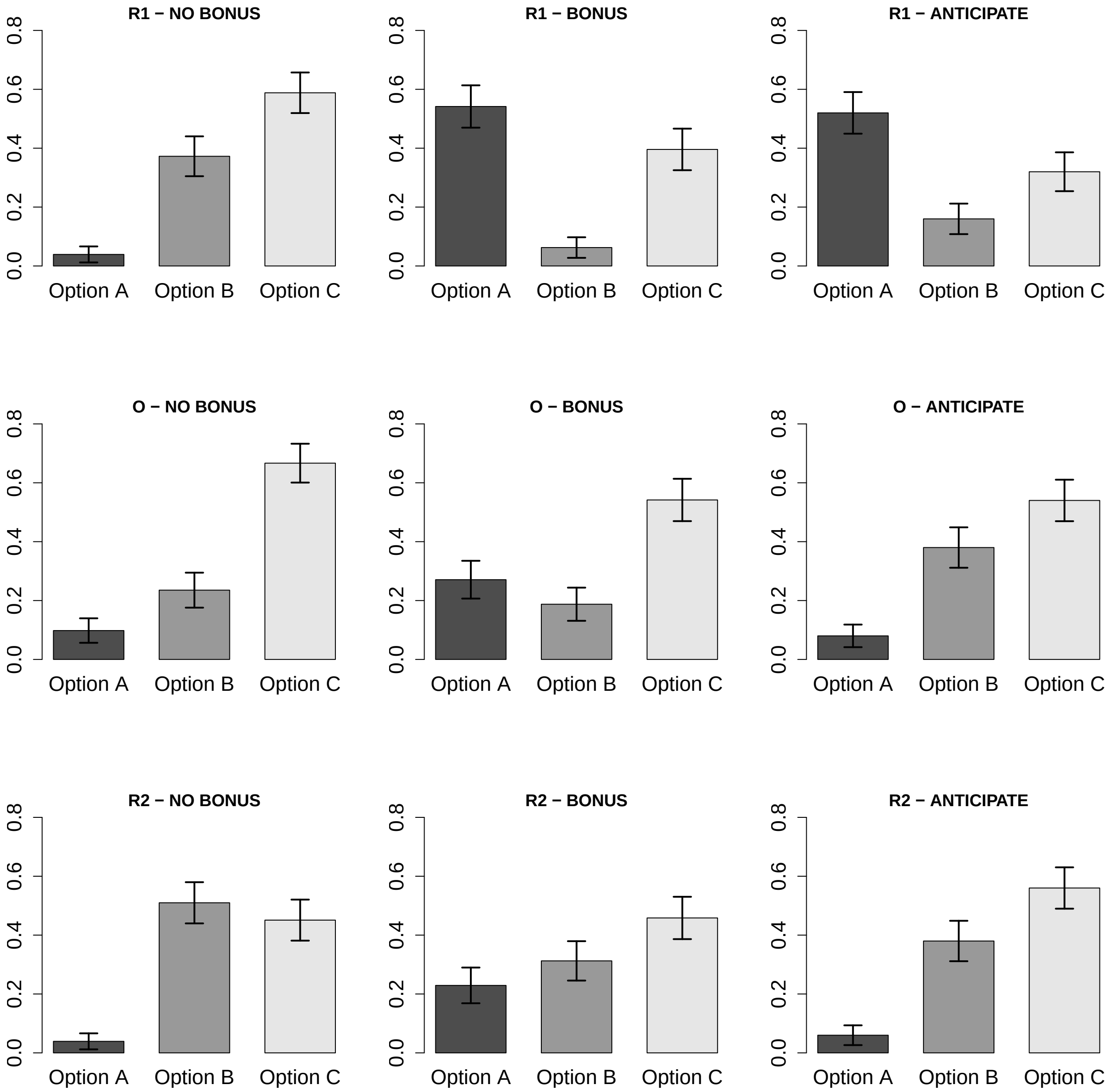
Lemma 8. *If advisers do not anticipate image costs, those with $r^* \in \mathcal{B}$ and $o^* \in \mathcal{B}$ prefer $\mathcal{BBB}$ among all possible plans of actions.*

Lemma 9. *If advisers do not anticipate image costs, all advisers with $r^* \in \mathcal{B}$ and $o^* \in \mathcal{B}$ choose either $\mathcal{BNN}$ or $\mathcal{NNN}$. The share of these advisers who prefer the former over the latter is given by $K(u(b))$.*

Proposition 3 then follows immediately from the above lemmas. □

Appendix C: Additional data

Figure C.1. Full distributions of advisers' actions (rows) over over treatments (columns)



Note: Bars depict standard errors.

Table C.1. Summary statistics for advisers' personal characteristics.

	NO BONUS		BONUS		ANTICIPATE		OVERALL		KW/χ^2-test
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	p-value
Age	24.82	8.002	23.208	5.411	25.02	7.38	24.37	7.055	0.110
Male	0.451	0.503	0.354	0.483	0.360	0.485	0.389	0.489	0.536
Region of origin									0.049
Europe/N. America/Australia/NZ	0.353	0.483	0.333	0.476	0.440	0.501	0.376	0.486	
Asia	0.608	0.493	0.646	0.483	0.420	0.499	0.557	0.498	
Other	0.039	0.196	0.021	0.144	0.140	0.351	0.067	0.251	
Degree									0.460
Bachelor	0.608	0.493	0.500	0.505	0.520	0.505	0.544	0.500	
Master	0.353	0.483	0.479	0.505	0.460	0.503	0.430	0.497	
PhD	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
Other postgraduate	0.000	0.000	0.021	0.144	0.000	0.000	0.007	0.082	
None	0.039	0.196	0.000	0.000	0.020	0.141	0.020	0.141	
Study subject									0.294
Economics/Business/Finance	0.216	0.415	0.375	0.489	0.360	0.485	0.315	0.466	
Other social sciences	0.353	0.483	0.229	0.425	0.300	0.463	0.295	0.458	
Psychology	0.059	0.238	0.021	0.144	0.080	0.274	0.054	0.226	
Public administration	0.039	0.196	0.063	0.245	0.020	0.141	0.040	0.197	
Math/Sciences/Engineering	0.157	0.367	0.083	0.279	0.020	0.142	0.087	0.283	
Arts or Humanities	0.157	0.367	0.146	0.357	0.140	0.351	0.148	0.356	
Other	0.020	0.140	0.083	0.279	0.080	0.274	0.060	0.239	
Monthly budget (in GBP)	606.3	450.7	640.0	563.8	909.4	559.9	718.9	540.4	0.088
Number of observations	51		48		50		149		

Note: The rightmost column provides p-values for the null of equality between the three treatments (Kruskwal-Wallis tests for the variables age and budget; χ^2-tests for the remaining categorical variables).

Appendix D: Experimental instructions

The following pages contain screenshots of instructions shown to on computer screens and on the information sheet about the investment options printed on paper. They are presented in the order as they were seen by the subjects in the experiment.

- Screen 1: Welcome stage and general instructions
- Screen 2a–2c: Explanation for R1. Three screens which explain the client's choice situation, the adviser's role and, if applicable, the bonus (Screen 2a), information about the upcoming decision situation (Screen 2b, only in ANTICIPATE), and the investment options (Screen 2c).
- Information on the investment options shown to advisers, printed on paper
- Screen 3: Instructions for giving the first recommendation R1
- Screen 4: Instructions for making the own choice O
- Screen 5: Instructions for giving the second recommendation R2
- Screen 6: Exit questionnaire

Information shown only in BONUS or ANTICIPATE is put in [ ]-brackets, information which is shown only in ANTICIPATE is put in [[ ]]-brackets.

Welcome to this experiment!

For participating in this experiment every one of you receives an amount of GBP 5.00.
During the experiment you can earn additional money depending on your decisions.
The whole experiment takes about 45 mins. You will be paid after all participants have finished.
So please take your time and pay attention when reading the instructions.

Please note that talking is not allowed during the experiment.
It is also not allowed to communicate using your mobile phones or other devices.

Please do not use the provided computer for anything else than this experiment.
In particular, you are not allowed to exit this program and/or switch to other functions of the computer.

Failure to comply with these instructions endangers the smooth running of the experiment and its scientific validity.
If you are caught to not comply, you may be excluded from this and future experiments and will not be paid.

Thank you for your understanding!

If you have any problems during the experiment, please keep quite and hold your hand out of the cubicle you are sitting in.
We will then come to you.

Screen 1: Welcome stage

General Information

Your role:

All subjects in the current experimental session are assigned the role of an **advisor**.
As an advisor, you will give a recommendation to a client.
These clients will be subjects in another experiment at the LSE's Behavioral Research Lab.

How it works

In this future experiment with clients, each of them has to choose one out of three options, A, B or C.
Here is what will be shown to the client:
"*Each option will earn different monetary payoffs.*
Option A presents a possibility to earn a high or a low payoff, depending on luck.
Option B adds the possibility to earn some amount between the high and low payoff, option C increases that possibility."

Clients however do NOT know more about this situation than the above text when they choose an option.
You, as an advisor, will soon learn what exactly these options are.
Afterwards, you have to recommend one option to a client.

Verification

You will have to write down your recommendation on paper and put it into an envelope. If you want, you can address the envelope to yourself.
At the end of this experiment, we will randomly choose one of the recommendations given here to be shown to a client.
If your recommendation is chosen to be shown to a client the following happens:

- We will read out loudly your cubical number (not the name) of that recommendation. You therefore know that you have been chosen.
- We will ask the client who will receive your recommendation to sign it.
- If you wrote your address on the envelope, we will mail you a copy of your recommendation signed by the client
- We will also mail you information of how you can retrive the receipt signed by that client from the lab's official record depository.
- The client will only see your written recommendation, not the envelope which potentially bears your name.

With this procedure, you can verify whether a client has actually gotten your advice, should your recommendation be drawn and you self-addressed the envelope.

[**Your bonus**

You receive a bonus of GBP 3.00 for recommending **Option A**
The bonus will be paid independently of whether your recommendation is chosen to be shown to a client.]

I understood. Please proceed.

Screen 2a: The clients' choice situation and information about the bonus.

General Information - Further Steps

A choice for your own and a second recommendation

After you have made a first recommendation to a client, there will be **two further steps**:

- First, you will have to choose one of the three options for yourself.
- Then, you will have to make another recommendation to a second client.
  (This client has the same information as the first client and will not have made a choice and will not have received advice before.)

As with your first recommendation, you will have to write down your own choice and the second recommendation on separate sheets of paper and put them in separate envelopes. We will then sample one of the envoples with own choices by you and other advisers in this session. As with the first recommendations we will also sample one of the envelopes with the second recommendations to a another clients in this session.

- The adviser whose own choice is sampled will have to take the chosen option and will be paid coordingly.
- The sampled second recommendation will be shown to a client.

Different to the first recommendation, you will NOT receive a bonus for any option you choose for yourself or any option which you recommend in the second recommendation.

I understood. Please proceed.

Screen 2b: Information about upcoming decisions (only shown in ANTICIPATE).

You will now learn precisely how a chosen option affects a client's payoffs in addition to the GBP 5.00 they get (as you will) for coming here.

A risky choice

You have to choose one out of the following three options to recommend to a client.
These options will be the same when you lafter first have to make a choice for your own and then make a second recommendation to another client.
This will determine the client's payoff as follows:

Option A

- Client rolls a six-sided die;
- For any number of the die: client flips a coin and earns GBP 20.00 when the coin shows "Heads"; or nothing when the coin shows "Tails".

Option B

- Client rolls a six-sided die;
- Die shows 1 or 2: client earns an amount of GBP 12.00;
- Die shows 3, 4, 5 or 6: client flips a coin and earns GBP 20.00 when the coin shows "Heads"; or nothing when the coin shows "Tails".

Option C

- Client rolls a six-sided die;
- Die shows 1 or 2: client earns an amount of GBP 12.00;
- Die shows 3 or 4: client earns an amount of GBP 8.00;
- Die shows 5 or 6: client flips a coin and earns GBP 20.00 when the coin shows "Heads"; or nothing when the coin shows "Tails".

Note: [Your bonus of GBP 3.00 which you get for recommending option A in the first recommendation is independent of a client's choice.]

[[While the options from which you have to choose for yourself and make a second recommendation will be **exactly the same as above** it will be a separate die roll (and, potentially, coin flip) for each recommendation and for the choice for yourself which determines the outcome.]]

Please look now at the paper instructions. It contains a summary of the above and a table which lists all possible outcomes.

Please study the table and examples carefully.
You will soon have to make a recommendation to the client. As said, the client knows nothing of the above.
If you are ready click "Continue" below.

Continue

Screens 2c: Information about the clients' investment option.

A risky choice

One of the following options must be chosen. Then the following happens:

Option A:

- Roll die: for every outcome, play the lottery.

Option B:

- Roll die: if it shows 1 or 2, one earns GBP 12.00 for sure;
- Roll die: if it shows 3, 4, 5 or 6, one has to play the lottery

Option C: receive a chance to roll the same six-sided die:

- Roll die: if it shows 1 or 2, one earns GBP 12.00 for sure;
- Roll die: if it shows 3 or 4, one earns GBP 8.00 for sure;
- Roll die: if it shows 5 or 6, one has to play the lottery

The lottery:

For the lottery one has to toss a coin. “Heads” then yields GBP 20.00, “Tails” nothing.

Each row of the table below represents a possible result of the die. The columns describe the possible consequences, depending on the chosen option.

Die equal to….	**Option A** is chosen	**Option B** is chosen	**Option C** is chosen
1 or 2	lottery: GBP 20 or 0	GBP 12	GBP 12
3 or 4	lottery: GBP 20 or 0	lottery: GBP 20 or 0	GBP 8
5 or 6	lottery: GBP 20 or 0	lottery: GBP 20 or 0	lottery: GBP 20 or 0

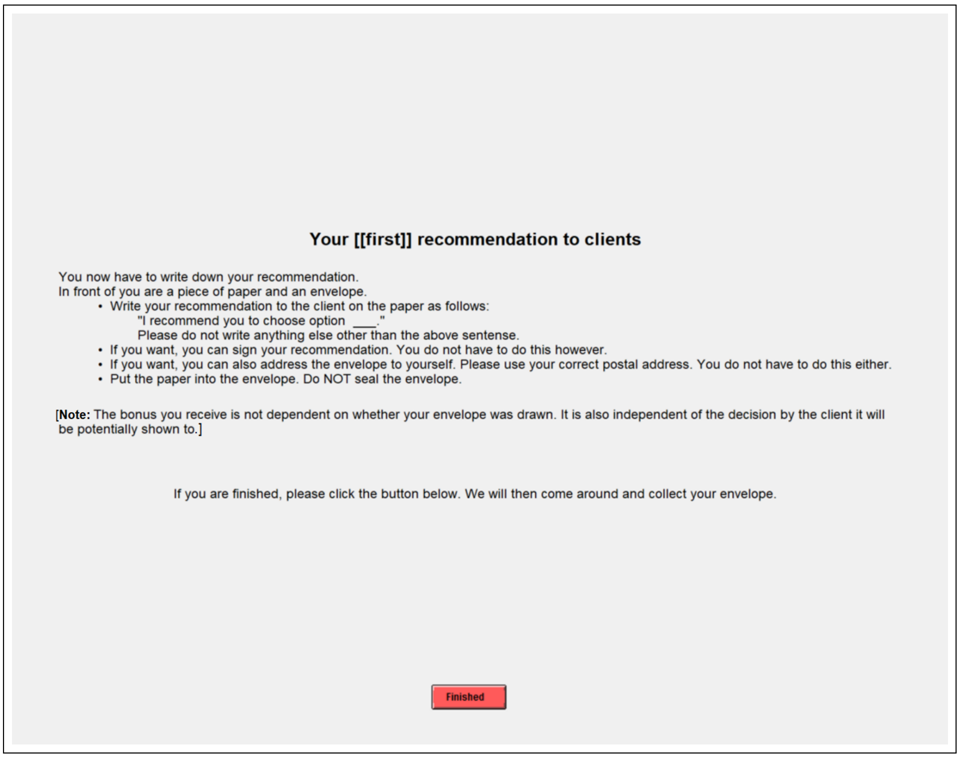
Example:

Suppose the die yielded 3: If option A or B was chosen before, one has to play the lottery. If option C was chosen, one would have gotten GBP 8.00 for sure instead.

Suppose the die yielded 1. If option B or C was chosen before, one gets GBP 12.00 for sure. If option A was chosen, one plays the lottery instead.

Suppose the die yielded 6. Independently of the chosen option one plays the lottery.

Information sheet shown to advisers
(It was placed face down on each adviser's table with the following print on its back:
"Information – do not turn until explicitly told so".)



Your [[first]] recommendation to clients

You now have to write down your recommendation.
In front of you are a piece of paper and an envelope.

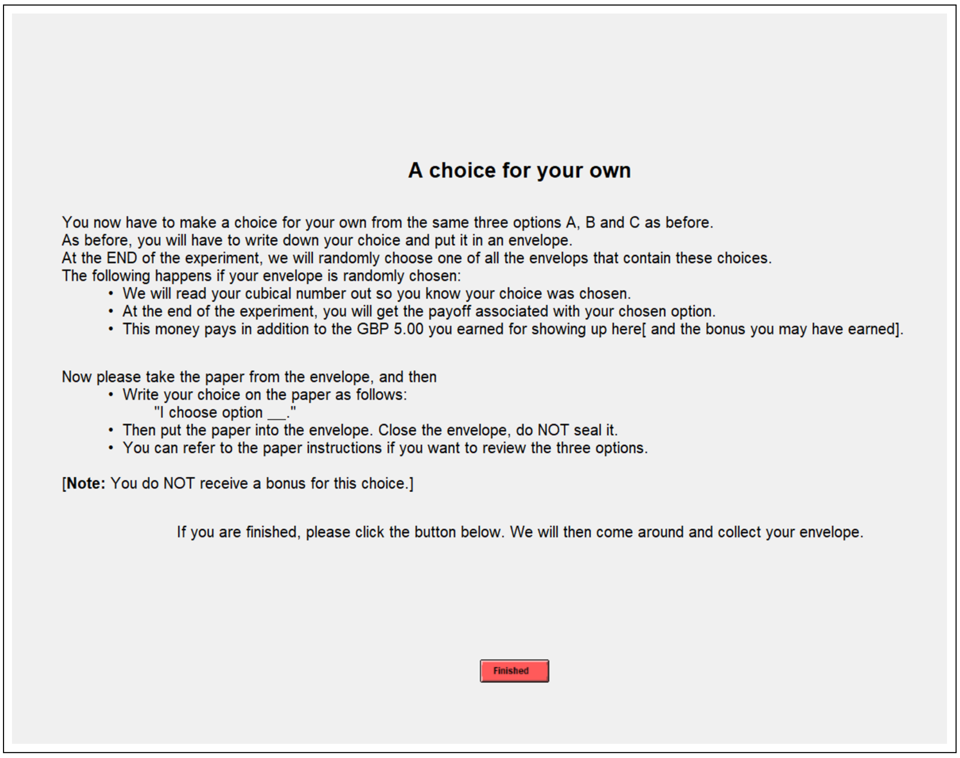
- Write your recommendation to the client on the paper as follows:
  "I recommend you to choose option ___."
  Please do not write anything else other than the above sentense.
- If you want, you can sign your recommendation. You do not have to do this however.
- If you want, you can also address the envelope to yourself. Please use your correct postal address. You do not have to do this either.
- Put the paper into the envelope. Do NOT seal the envelope.

[**Note:** The bonus you receive is not dependent on whether your envelope was drawn. It is also independent of the decision by the client it will be potentially shown to.]

If you are finished, please click the button below. We will then come around and collect your envelope.

Finished

Screens 3: Instructions for giving the first recommendation R1.



A choice for your own

You now have to make a choice for your own from the same three options A, B and C as before.
As before, you will have to write down your choice and put it in an envelope.
At the END of the experiment, we will randomly choose one of all the envelops that contain these choices.
The following happens if your envelope is randomly chosen:

- We will read your cubical number out so you know your choice was chosen.
- At the end of the experiment, you will get the payoff associated with your chosen option.
- This money pays in addition to the GBP 5.00 you earned for showing up here[and the bonus you may have earned].

Now please take the paper from the envelope, and then

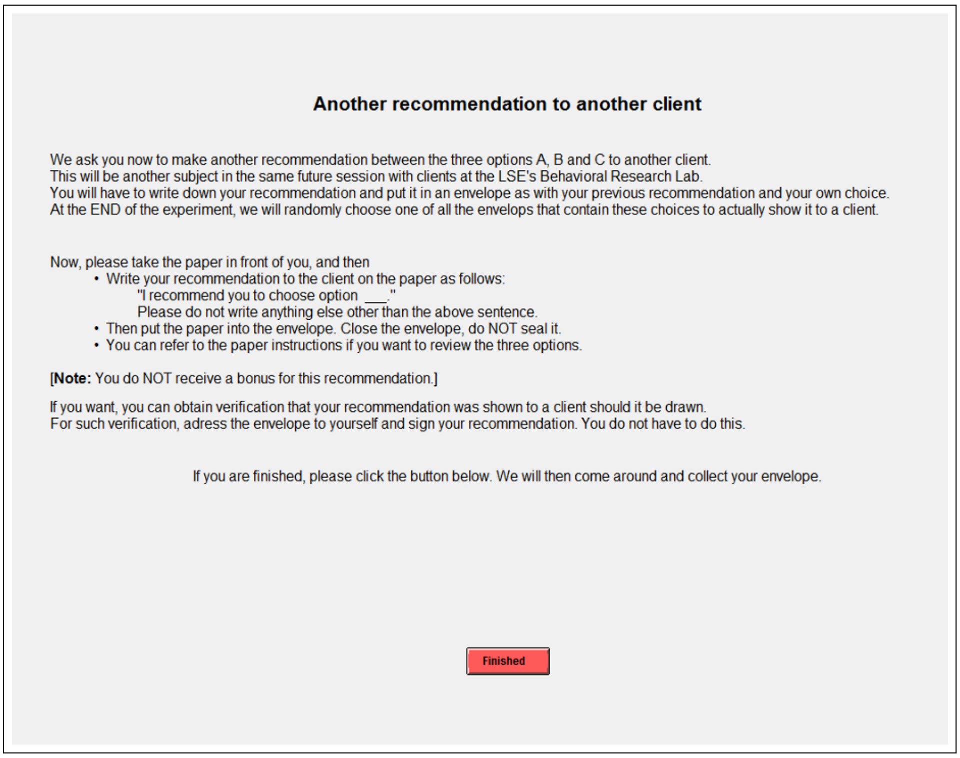
- Write your choice on the paper as follows:
  "I choose option __."
- Then put the paper into the envelope. Close the envelope, do NOT seal it.
- You can refer to the paper instructions if you want to review the three options.

[**Note:** You do NOT receive a bonus for this choice.]

If you are finished, please click the button below. We will then come around and collect your envelope.

Finished

Screen 4: Instructions for making the own choice O



Another recommendation to another client

We ask you now to make another recommendation between the three options A, B and C to another client.
This will be another subject in the same future session with clients at the LSE's Behavioral Research Lab.
You will have to write down your recommendation and put it in an envelope as with your previous recommendation and your own choice.
At the END of the experiment, we will randomly choose one of all the envelops that contain these choices to actually show it to a client.

Now, please take the paper in front of you, and then

- Write your recommendation to the client on the paper as follows:
  "I recommend you to choose option ___."
  Please do not write anything else other than the above sentence.
- Then put the paper into the envelope. Close the envelope, do NOT seal it.
- You can refer to the paper instructions if you want to review the three options.

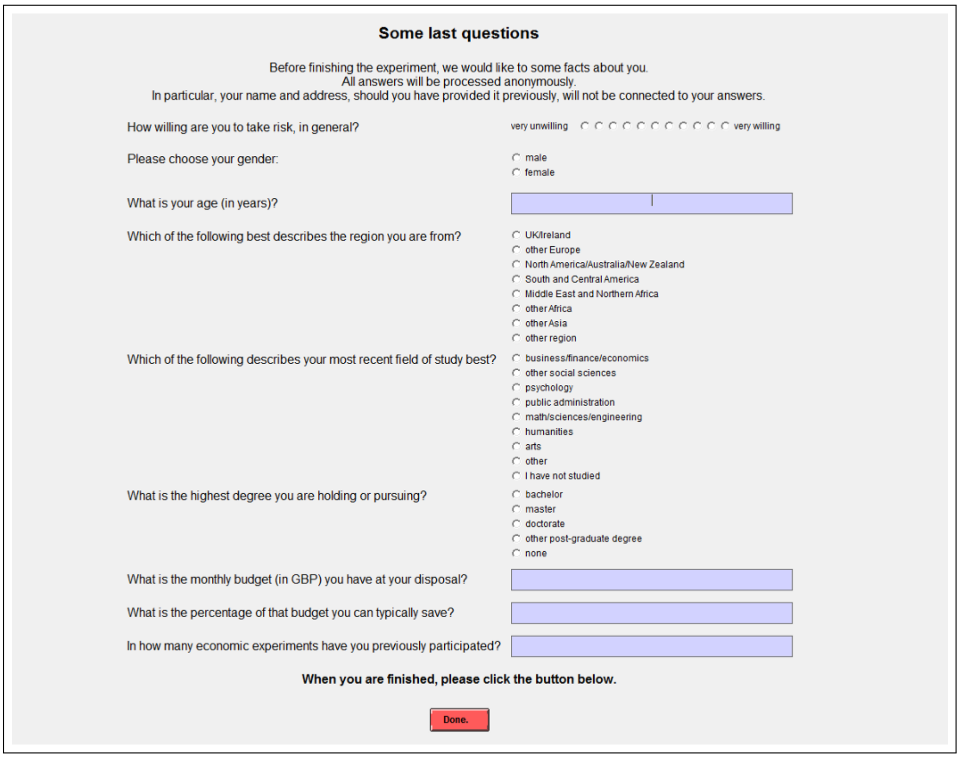
[**Note:** You do NOT receive a bonus for this recommendation.]

If you want, you can obtain verification that your recommendation was shown to a client should it be drawn.
For such verification, adress the envelope to yourself and sign your recommendation. You do not have to do this.

If you are finished, please click the button below. We will then come around and collect your envelope.

Finished

Screen 5: Instructions for giving the second recommendation R2



Some last questions

Before finishing the experiment, we would like to some facts about you.
All answers will be processed anonymously.
In particular, your name and address, should you have provided it previously, will not be connected to your answers.

How willing are you to take risk, in general?	very unwilling ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ very willing
Please choose your gender:	○ male ○ female
What is your age (in years)?	
Which of the following best describes the region you are from?	○ UK/Ireland ○ other Europe ○ North America/Australia/New Zealand ○ South and Central America ○ Middle East and Northern Africa ○ other Africa ○ other Asia ○ other region
Which of the following describes your most recent field of study best?	○ business/finance/economics ○ other social sciences ○ psychology ○ public administration ○ math/sciences/engineering ○ humanities ○ arts ○ other ○ I have not studied
What is the highest degree you are holding or pursuing?	○ bachelor ○ master ○ doctorate ○ other post-graduate degree ○ none
What is the monthly budget (in GBP) you have at your disposal?	
What is the percentage of that budget you can typically save?	
In how many economic experiments have you previously participated?	

When you are finished, please click the button below.

Done.

Screen 6: Exit questionnaire