Persistent Bias in Advice-Giving

Zhuoqiong (Charlie) Chen  Tobias Gesche
HIT Shenzhen  ETH Zurich

July 2018

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 which is only attractive for risk-seeking individuals, relative to its alternatives. Afterwards, they 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 does not only bias initial recommendations, but also subsequent actions: Advisers who were offered the bonus only for their first recommendation chose the risky option and recommend it a second time up to six times more often than advisers in a control group who were never offered a bonus. In a further treatment, 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 whole sequence of actions does not decrease the bias in the initial recommendation. However, it is effective in eliminating the bonus’ persistent effect on the subsequent own choices and second recommendations. These results are consistent with a theory we present and which combines biased advice-giving and advisers’ (self-)image concerns of appearing incorruptible.

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; Linnainmaa et al., 2017). Such conflicts can even be generated more subtly through unconditional gifts (Malmendier and Schmidt, 2017) – their consequences are however vast: For retirement investment in the US alone, which is just 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 17 billion dollars 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 stakes might even be higher, albeit harder to quantify. In light of these economic and ethical problems and the fact that disclosure often does not help, removing the cause of the conflict of interest seems appealing.\footnote{There is now numerous evidence that disclosing, as opposed to removing, conflicts of interest of experts may not only be ineffective, but also 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 underlying these effects; complementing economic accounts are presented in Li and Madarasz (2008), Inderst and Ottaviani (2012), and Gesche (2016).} In fact, policies which aim at removing and banning adverse incentives for advisers have been brought up or are being considered in various jurisdictions.\footnote{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).} This paper explores how such incentives can, even after their removal, affect adviser behavior.

In an experiment, we document that many advisers’ recommendations are biased and that this bias persists, after the underlying conflict of interest has disappeared. Subjects who acted as advisers were paid a bonus when they recommended a risky option which only a risk-seeking person would prefer, relative to its alternatives. Clients received this advice and did not know the payoffs and the distribution for this option or its alternatives. Advisers had this information. About half of the advisers recommended the risky option when they were offered a bonus whereas only a small minority did so in a control condition without a bonus. Advisers then learned that they would also have to choose for themselves, from the same set of options but without any bonus. 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 advisers who were previously offered the bonus were more than thrice as likely to choose the risky option for themselves than those in a control condition where no bonus was ever paid. The removed bonus also affected second recommendations: Advisers who had previously been exposed to the bonus were six times more likely to recommend the risky option to another client than advisers in the control condition.

These findings are consistent with a simple theory we present. Its underlying reasoning is based on two main notions. The first is that being influenced by a bonus, i.e., giving biased advice, is deemed immoral. Consistent with this, almost half 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. In line with such a mechanism, we estimate that around 40% of those advisers whose initial advice was biased recommended the risky option again.

Our results also allow to explore how advisers determine appropriate advice. For those advisers whose initial advice was corrupted by the bonus we estimate that a sizable fraction – again around 40% – also chose the risky option for themselves. This finding suggests that advisers linked their own choices to what they consider appropriate advice and then had to choose accordingly to avoid signaling their corruptibility. It speaks against an alternative reasoning where they could have self-servingly assumed that their clients are risk-seeking. If this were the case, there would not have been a negative signal if they had chosen a different option for themselves than the one they had previously recommended in order to earn the bonus.

To further investigate the mechanisms which underly the persistent biases we document, we also conducted an additional treatment. Similar to the previous treatment with a bonus, advisers 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, first recommendations were as biased as when a bonus was paid and the following own choices and recommendations were not announced before-hand. 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 compared to the condition in which no bonus was ever paid. In line with a simple extension of our theory to a dynamic decision problem, these findings point towards the relevance of image concerns which are only backward-looking. Our results also rule out alternative explanations like anchoring or inertia which could have created the differences between the other two treatments. Finally, these findings also demonstrate the potential of ex-ante considerations to affect moral reasoning, change behavior, and 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 results. Foerster et al. (2017) use observational data on about 6,000 Canadian financial advisers and more than 580,000 of their clients. They show that not clients’ personal characteristics but simple fixed effects for the individual advisers explain most of the variation in how risky the clients’ investment portfolios are. Using the same data set, Linnainmaa et al. (2016) report that recommendations to clients are also reflected by the choices which these advisers make for themselves: They chose the same return-chasing and actively managed funds which their respective clients held. These advisers did so, even though these investments were riskier than others and, net of fees, performed worse than the market average. Our work explains how commissions (which are typically paid for selling such funds) can cause these patterns. We use an experimental approach so that exposure to the bonus can be varied exogenously. This allows us to exclude an alternative theory in which advisers self-select into compensation schemes which reward their personal preferences. Rather, we find support for the notion that such rewards can, in a persistent and causal way, bias advisers’ recommendations and their own choices.

Closely related to our findings are also those of an experiment by Gneezy et al. (2016). For a one-shot recommendation, they report a relatively low bias when advisers first saw a set of risky options and then learned about a bonus to recommend a specific one. When this order is reversed, i.e., when advisers first learned about the bonus before they could see the option, this bias increases, supposingly because this order allows a biased initial assessment of the options. This finding and this explanation are in line with our results and the mechanism we propose. More generally, our findings also connect to a related literature which shows that role-induced dispositions lead people to align their judgment. In a classic study, Festinger and Carlsmith (1959) show that paying subjects to report favorably about an unpleasant task to others improves their subsequent evaluation of this task, relative to when they were
not paid. Loewenstein et al. (1993) report on an experiment in which subjects acted in the fictitious role of plaintiff or defendant in a legal case. These roles shifted what the subjects considered a fair settlement value towards the roles’ respective positions. In the same experimental setting, Babcock et al. (1995) report that subjects find it much harder to agree on a settlement value when they knew their role before learning about the case’s details than vice versa. Similarly, Konow (2000) finds a systematic shift for entitlements of a collectively generated surplus, depending on the roles a subject had in a subsequent dictator game in which this surplus was split. That such induced judgments can carry over to own actions in a harmful way, i.e., that people self-deceive (Trivers, 2011), is demonstrated in a recent study by Schwardmann and van der Weele (2016). They show that when tasked to convince others of their own ability, subjects overstate their own ability in subsequent private self-assessments, even when doing so is costly to themselves. These patterns can be explained by a desire to minimize cognitive dissonance (Festinger, 1957), which arises when one’s expressed opinion does not reflect one’s actual judgments. \(^{3}\)

Our work takes up on these insights in the context of advice-giving. We demonstrate how such reasoning can diminish the effectiveness of removing conflicts of interest. In the theory we lay out, consistency with biased advice in own choices and repeated advice is instrumental in preserving a positive image. A similar logic underlies the findings by Falk and Zimmermann (2017a,b). They report that subjects forfeit opportunities to improve their accuracy in an estimation task in order to be consistent and thereby signal ability to a principal or themselves. Mijović-Prelec and Prelec (2010) report a similar pattern of costly self-deception and how it can be derived through a self-signaling mechanism. We show how such harmful consistency can emerge through conflicts of interest and how it persistently affects the role of advisers as information providers. In doing so, we relate to the literature on the moral underpinnings of sender behavior in strategic communication (e.g. Gneezy, 2005; Sutter, 2009; Rode, 2010; Inderst et al., 2017). Instead of a sender-receiver game, we consider the related but different situation of advice-giving.\(^{4}\) We also connect to the recent literature on the adverse effects of bonus payments (Christoffersen et al., 2013; Bénabou and Tirole, 2016) and how the resulting conflicts of interests shape the self-perception and attitudes of those exposed to them, for example in the context of financial services (Zingales, 2015; Cohn et al., 2014). However, our findings come from a neutral framing and also relate to biased advice more generally.

\(^{3}\)Kunda (1992) discusses how cognitive dissonance and (motivated) self-perception relate. For economic models of cognitive dissonance, see Akerlof and Dickens (1982) and Rabin (1994).

\(^{4}\)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.
Finally, we also contribute 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; Falk and Tirole, 2018). Although such image concerns can refer to both, social and self-image, the latter alone can steer moral behavior. This applies, for example, to non-maximal lying in order to uphold the illusion of being honest (Mazar et al., 2008), inflicting less harm on others while seeing oneself on a video screen (Falk, 2017), or paying more under a pay-what-you-want scheme than under fixed prices to avoid appearing greedy to oneself (Gneezy et al., 2012). We expand this literature by linking image concerns to moral reasoning in the context of advice-giving. As a key theoretical result, we show how this combination can cause persistence of the biases which conflicts of interest generate. This predicted persistence is then supported by the results of our experiment in which social, but not self-image concerns were minimized.

A related branch of this literature, recently summarized by Gino et al. (2017), shows that often, information is not processed in an objective manner if this threatens a person's self-image. Rather, people act as "motivated Bayesians" who employ uncertainty and ambiguity in a self-serving way (Dana et al., 2007; Haisley and Weber, 2010; Exley, 2016). Importantly, this includes the formation of beliefs about others and others’ preferences to accommodate own selfish actions (Di Tella and Pérez-Truglia, 2015). Our findings cannot be explained solely by self-serving beliefs about others, here clients’ risk preferences. They rather point towards a link between biased advice and own choices with lasting consequences for both.⁵

3 A psychological mechanism

In this section, we describe how a preference to appear as an unbiased adviser can lead to a repeated bias in further advice and in own choices. The underlying mechanism can also be derived in a formal model (see Appendix A). It is based on an adviser ("he") who advises a client ("she") and who 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 three elements: 1) utility derived from expected monetary payoffs, 2) psychological or material costs of not giving appropriate advice, and 3) diagnostic dis-utility of learning from one’s current actions that previous advice was biased.

---

⁵This also connects our’s to earlier research which shows that people base their estimates of others’ preferences, in particular risk preferences, on their own (Mullen et al., 1985; Faro and Rottenstreich, 2006).
While the first element is standard, the second reflects advisers’ uneasiness to recommend something they do not consider appropriate. For example, an adviser might think that a recommendation for a particular choice is suited for 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 one has not acted in the client’s best interest.\(^6\) However, predicting others’ preferences is inherently difficult. This applies in particular for risk preferences (Hsee and Weber, 1997; Eckel and Grossman, 2008; Harrison et al., 2013). Even trained financial advisers who receive information about the clients’ characteristics and who have no conflict of interest have difficulties in predicting their client’s risk preferences (Roth and Voskort, 2014). When there is a conflict of interest this uncertainty can be instrumentalised in a self-serving manner: Advisers may form a self-serving belief about their clients’ preferences such that this belief is compatible with their biased advice.

There are, however, limits to such self-serving beliefs. 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).\(^7\) Some advisers might also determine what they consider appropriate advice by the answer to the question ”What would I choose if I were in the client’s position?” Appropriate advice may also be determined via a reasoning which is independent of an adviser’s own preferences. For example, beliefs over clients’ preferences can be formed in self-serving manner such that it facilitates to recommend an option to which an external incentive point. Our theory and the experiment we conduct allow to explore the different implications which these lines of reasoning have. However, for the principal mechanism of repeated biased advice we explore, it only matters that costs of recommending something inappropriate exist, irrespective of how (in)appropriate advice is determined.

The third factor which matters for advisers is diagnostic (dis-)utility which they experience when their actions reveal that they have given biased advice, for example through a self-signaling mechanism. In contrast to the costs of giving inappropriate advice, this dis-utility only occurs to an adviser after

\(^6\)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.

\(^7\)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.
he has biased his advice, at the point when his later actions indicate exactly this fact to him. This can be captured by a dual-self model in which the "diagnostic self" of an adviser learns ex post about the adviser’s motive for giving advice, e.g. whether prior advice was biased or not. 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 which are incompatible with this notion. Dual-self models have been used previously to explore how people infer about themselves, in particular, their morality (e.g. Bodner and Prelec, 2003; Bénabou and Tirole, 2011; Grossman and van der Weele, 2017). Here, we use such an approach to describe the trade-off between keeping self-serving beliefs about one’s own motives and taking contradictory actions.8

These three components together then have implications for how and, most importantly, for how long conflicts of interest can affect advisers’ actions. To see this, consider an adviser whose recommendation was corrupted: His cost of giving biased advice, i.e., the cost of recommending something inappropriate, is thus smaller than the (material) benefit he gets from giving such advice. If the adviser is also sufficiently concerned about his self-image, he then needs to continue to give the same biased advice again when the conflict of interest has disappeared. The reason is that in order 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 disappeared signals the opposite.

As mentioned above, an adviser’s own preference can also determine what he thinks should be recommended to a client. In this case, an adviser whose advice was corrupted and who has then to choose for himself is also put on the spot: If he does not choose as he has recommended, he also signals his previous corruptibility. Accordingly, if advisers’ image concerns are sufficiently high, they effectively bias their own choices. In summary, a behavioral trait which generally seems to be desirable – a preference to be perceived as unbiased – can lead to a persistent bias in the context of advice-giving. In addition, it can have a lasting effect on advisers’ own choices.

The preceding reasoning treats adviser behavior as a series of static decisions. Backward-looking image concerns can then create a spill-over from having given biased advice after the bonus’ removal. This matches several realistic and important situations. For example, financial advisers may have given advice under a conflict of interest for a long time before a new regulation which removes such conflicts

8In the context of this paper, we mainly talk about image concerns as self-image concerns because this is, as we will argue, most consistent with our experimental setup. However, the main mechanism we propose and which underlies a persistent bias can also be triggered by social image concerns, i.e., when an outside observer’s perception matters, or by social and self-image concerns together.
is first discussed and then, eventually, introduced. However, there are also scenarios where advisers do know that future actions will follow and that incentives to bias advice are only one-off. A typical situation in this category is when advisers have a temporary incentive to sell a certain good or service, for example because stocks are expiring or vacancies are left unused. Such settings in which upcoming actions and the bonus’ temporary nature can be anticipated transform the decision of whether to bias the initial advice. Instead of being a isolated decision, it becomes a crucial part of a dynamic decision problem.

The above framework used to analyze adviser behavior can be extended to capture such dynamic considerations. In principle, there are two possibilities. They vary by whether advisers factor in backward-looking image costs when they initially form a plan of actions, that is, when they initially decide whether to bias advice. If they do so, the anticipation of image costs creates an additional cost of initially giving biased advice. This is because such advice forces advisers 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 prior corruptibility. If advisers do not factor in backward-looking image concerns, deviating from an initially biased advice is perceived to be less costly. That is, anticipation can cause advisers to plan to behave in a "colder" fashion, without regards to future image concerns, and thereby dampen the persistent effects of one-off incentives to bias advice.

4 Experimental design and procedures

To explore empirically whether a bonus to bias advice has a persistent effect on own choices and future recommendations, and how such persistent effect depends on advisers’ anticipation, we conducted a controlled experiment. At the beginning of the experiment subjects were allocated to computer terminals in cubicles where instructions were shown to them on screen. Subjects who acted as advisers were informed that they would get GBP 5.00 as a show-up fee for participating in the experiment and that there would be further possibilities to earn money. They were then informed that they would act as advisers for clients in a future experimental session and that these clients would be drawn from the same pool of subjects and receive the same show-up fee.

Advisers had to recommend one of three risky choices, referred to as option $A$, $B$, and $C$ to their clients. They were informed that option $A$’s payoff would be either high or low, depending on luck, that option $B$ would also allow to get an intermediate payoff, and that option $C$ increases the probability for getting an intermediate payoff. Advisers knew that this information would also be given to the
clients but that clients would neither know the options’ payoffs nor the respective probabilities. They, as advisers, would soon learn these parameters before they had to make a recommendation. The advisers’ superior information was then given to them on a paper sheet which explained the three investment options in detail (for a copy of this sheet and the computer interface see Appendix D). The text on the sheet explained the following procedure of how an option’s payment was determined: After an option was chosen, a six-sided die would be rolled. Depending on the chosen option, this would then yield either a safe payment or a lottery. This lottery was described as a (fair) coin toss with “Heads” yielding GBP 20 and “Tails” nothing. Table 1 summarizes this. It was also printed on the paper instructions, together with a text and examples which explained this procedure in detail. Throughout the experiment, advisers could keep this paper and refer back to it when needed.

Note that a choice among the options, i.e., among the three compound lotteries they represent, allows to categorize the underlying risk preferences: If one compares option A and B, only a person who is willing to give up a safe payment of 12 to play a gamble with an expected payment of 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 10 for a safe payment of 8. Thus, only a risk-averse individual should choose option C. Accordingly, option B is chosen by a person who is approximately risk-neutral, i.e., neither sufficiently risk-averse nor sufficiently risk-seeking. Reflecting this ordering, we will henceforth refer to option A/B/C as the risky/neutral/safe option, respectively.9

<table>
<thead>
<tr>
<th>Die equal to:</th>
<th>Option A</th>
<th>Option B</th>
<th>Option C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 or 2</td>
<td>lottery: GBP 20 or 0</td>
<td>safe payment: GBP 12</td>
<td>safe payment: GBP 12</td>
</tr>
<tr>
<td>3 or 4</td>
<td>lottery: GBP 20 or 0</td>
<td>lottery: GBP 20 or 0</td>
<td>safe payment: GBP 8</td>
</tr>
<tr>
<td>5 or 6</td>
<td>lottery: GBP 20 or 0</td>
<td>lottery: GBP 20 or 0</td>
<td>lottery: GBP 20 or 0</td>
</tr>
</tbody>
</table>

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

---

9 This choice between possible sub-lotteries within a compound lottery is essentially a stripped-down version of tasks used previously by Hsee and Weber (1997) and Holt and Laury (2002). For example, Holt and Laury (2002) let subjects choose ten times among pairs of lotteries. Across the ten choices, each pair’s second lottery becomes increasingly risky. One of the ten choices is then randomly picked to be implemented. This allows to interpret the switching point between the first and second lottery as an indicator of risk preferences. We have essentially two such switching points (between A and B when the die equals 1 or 2 and between B and C when the die equals 3 or 4) which allow the categorization along risk-seeking/neutral/averse preferences.
After these initial explanations, the experiment proceed along five steps:

**Step 1 – First recommendation R1**: After having studied the instructions and choice situations, advisers 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 with the adviser’s cubicle number printed on it into an envelope and close it. The envelope was then collected by an experimenter and put into a box. Before they made their recommendations, advisers were told that at the end of the experiment, one of the envelopes would be randomly drawn from the box and the letter in it would be shown to a client who would then have to choose an option.

**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. Advisers were previously not informed about this step. The procedure was the same as for issuing advice: 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 it in a separate box. Again, they were informed that at the end of the experiment, one of the envelopes would be chosen randomly. The adviser who made the sampled choice then had to play the chosen option. Ex-ante, the choice situation and its implementation probability were thus the same as the one for which they had previously issued advice in R1, except that the choice was for themselves.

**Step 3 – Second recommendation R2**: After advisers had made their own choice O, they were asked to make a second recommendation. Again, this was not announced beforehand. The procedure was exactly the same as for R1, including the collection of envelopes in a separate box and announcing in advance that one would be sampled from it. Advisers were also informed that their second advice, if it was sampled, would be shown to another client who would not have received any previous advice. The decision situation thus mirrored exactly the first advice in R1.

**Step 4 – Questionnaire**: When all recommendations from R2 were collected, advisers had to fill out a short on-screen questionnaire which elicited personal information. It also included a short question on advisers’ general willingness to take risk.

**Step 5 – Payoff**: At the end of the experiment one envelope was sampled from each of the boxes for R1, O, and R2. The corresponding cubicle number (but neither the recommendation nor the choice by the subject) was read aloud so that the respective subject knew whether his or her envelope was sampled. The sampling procedure at the end of the experiment was announced and explained before advisers made their respective recommendations and choices. Subjects were then called, one by one, to
the laboratory’s exit where they were paid privately. In each session, the subject whose own choice was chosen to be implemented also rolled a die and, if necessary, also tossed a coin to determine the chosen option’s payoff. The subject was then paid accordingly.

**Treatments NO BONUS and BONUS:** The above describes the experimental procedure in our baseline condition to which we will refer to as NO BONUS. Our first experimental manipulation was to offer some advisers a bonus for recommending the risky option \( A \) in R1. We will refer to this treatment as BONUS. After having been informed that they had to give advice but before seeing the sheet with the detailed information about the investment options, advisers in BONUS were informed that they would get a bonus of GBP 3 for recommending option \( A \). This bonus was only paid for subject’s first recommendation R1, together with other earnings at the end of the experiment. For the advisers in BONUS, it was clearly stated on screens which explained the O and R2 tasks that there would not be any additional bonus for choosing or re-recommending option \( A \) in these tasks. This one-off bonus in R1 is the only difference between BONUS and NO BONUS.\(^{10}\)

**Treatment ANTICIPATE:** To investigate the effect which advisers’ knowledge of upcoming actions had and to rule out alternative explanations, we conducted a third, additional treatment. It largely resembles BONUS, i.e., it promised a bonus of GBP 3 if an adviser recommended option \( A \) in R1. Also as in BONUS, this bonus was not paid for own choices in O and second recommendations in R2. In contrast to BONUS however, advisers in ANTICIPATE were told before-hand not only about the bonus in R1, but also about O and R2 and that the bonus would be removed for these stages. They were informed about this on a separate screen which appeared after advisers had learned about the general setup of the advice-giving situation and the bonus but before they were actually asked to make their first recommendation. This information about the upcoming decisions in O and R2 and the bonus’ one-off nature is the only difference between ANTICIPATE and BONUS.

**Verifiability:** In order to ensure that advisers believed that a recommendation, if randomly chosen to be shown to a client, would be actually seen by the client we used the following procedure: We allowed advisers to voluntarily sign their recommendations and address the envelopes to themselves. Advisers were explained that if their recommendation was chosen to be shown to a client, the sheet

\(^{10}\)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.
would be signed by the respective client. In case that the corresponding adviser had provided us with his or her address, this subject would then get a copy of the signed recommendation by post. Subjects were informed of this option before they made their first recommendation and reminded of it before the second. It was also announced that the sampled recommendation’s cubicle number (but not the recommendation itself) would be announced at the end of this experiment. In this way, advisers knew that experimenters were pre-committed to actually show clients the sampled advice letters.

**General procedures:** Throughout the experiment, we enforced a strict no communication policy. We conducted eleven sessions, each with 11 to 17, in total 149, subjects acting as advisers. Advisers earned on average GBP 6.89 (around USD 9.50 when the experiments took place) while 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 descriptive statistics). The experimental sessions were conducted in late January 2016 (treatments NO BONUS and BONUS) and April 2018 (treatment ANTICIPATE) at the London School of Economics’s Behavioural Research Lab with subjects from its pool. Before the experiments, the study’s principal design and research questions were submitted as a part of the (successfully completed) procedure to obtain approval from the school’s research ethics committee. The experimental interface was implemented using zTree (Fischbacher, 2007). In the week after the adviser sessions, we invited additional subjects from the same pool for additional sessions. In these sessions, these subjects acted as clients and each received one of the sampled recommendations from the previous adviser sessions. After reading the recommendation, clients made their choices and were then paid accordingly. Advisers knew about this structure. In particular, they knew that each client and, if applicable, advisers themselves would each roll a die to determine a chosen option’s payoff so that payments were determined in an identical but independent manner. In this paper, we focus on advisers and their actions.\textsuperscript{11}

5 Predictions

In this section, we derive predictions for our experiment. They are based on the assumptions described in Section 3, thus on advisers maximizing their overall utility from direct pecuniary payoffs, costs of giving inappropriate advice, and image concerns of being perceived of having issued biased advice. Given the bonus’ form, we make the predictions with regards to how often the risky option $A$ is recommended.

\textsuperscript{11} Given that we have six relevant conditions (recommendations from R1 vs. R2 and BONUS vs. NO BONUS vs. ANTICIPATE) and only 22 client observations which are, due to our random sampling procedure, not balanced across these conditions, there is not much analysis which can be done due to limited statistical power.
and chosen. We proceed as follows: We first derive predictions regarding the comparison of BONUS and NO BONUS – our main treatments – over the three decisions situations (Predictions 1 through 3). Then, we derive how behavior changes in these situations when upcoming choices can be anticipated in ANTICIPATE and how this depends on whether advisers factor in upcoming image costs (Predictions 4a and 4b). All these predictions can also be derived within our formal model which can be found in Appendix A.

5.1 Predictions for the treatments BONUS vs. NO BONUS

In NO BONUS, there is no pecuniary gain of issuing any specific recommendation. Since this is the first action an adviser takes, it does not have signaling value with regards to past behavior. Therefore, image concerns do not matter. Absent other motives, only the costs of issuing inappropriate advice remain. Thus, only advisers who think that option $A$ is actually a good recommendation recommend it.

In the BONUS treatment, advisers are paid for recommending option $A$. In addition to those who actually think that this option is appropriate, some advisers might be induced to recommend it in order to earn the bonus although they do not consider option $A$ appropriate. This happens when the costs of giving inappropriate advice are low, relative to the pecuniary utility associated with the bonus. 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, costs of giving inappropriate advice play no role. In NO BONUS, there are also no image concerns of having issued biased advice before. Accordingly, advisers in this condition choose the option which they prefer (which might coincide with what they consider appropriate and thus have previously recommended, see below).

In the BONUS treatment this reasoning for advisers’ own choices does not go through. As previous advice in R1 could have been corrupted by the bonus, there is also a concern of being perceived or perceiving oneself as a biased adviser. This is relevant when appropriate recommendations relate to one’s own preferences, for example, when the appropriate advice is determined by the answer to the question ”What would I do if I were in the client’s situation”. The consequence of such reasoning is that for an unbiased adviser, his own choice in O and his previous recommendation in R1 should coincide. In the presence of image concerns, this has further implications for advisers who have initially recommended option $A$ in R1 just because of the bonus. When they now choose differently they signal that their initial advice was corrupted. Biased advisers can avoid this negative signal if they also choose
option \(A\) for themselves. This however leads to a loss in expected pecuniary utility as they choose option \(A\) instead of their truly preferred non-\(A\) choice. They therefore do so if the image costs are high, relative to this loss.\(^{12}\) In addition, those advisers who would have chosen option \(A\) anyhow because they truly prefer it, as in NO BONUS, choose \(A\) for themselves. Assuming that there is such reasoning which relates own choices to appropriate advice and that image costs are high enough for some advisers, we get the following prediction:

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

If the above prediction were wrong, this could have two reasons. The first is that advisers determine appropriate recommendations independently of what their own preferred choices are. 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 \(A\) without incurring the costs of giving inappropriate advice. Such a self-serving belief about the client’s preferences does not need to relate to their own preferences. In this case, choosing differently for themselves and for their client would not send a signal that the previous advice was biased which then yields the following, alternative prediction:

**Prediction 2b.** *In O, if advisers determine what appropriate advice is independently of their own preferences, 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. Absent such concerns, the initial, biased recommendation should not affect own choices. By the same reasoning, image costs should then also not affect the recommendations in R2. However, if image costs do matter, they predict treatment differences in this stage, irrespective of how appropriate advice is determined:

For recommendations in R2, an adviser’s own pecuniary utility is unaffected by what is recommended as in this stage there is no incentive to bias advice. In NO BONUS, there was also no such incentive in any previous stages so that image concerns of having given biased advice cannot play any role either. Accordingly, only the costs of giving inappropriate advice matter. Option \(A\) is then only recommended by those who have already recommended it in R1 because they genuinely consider this option appropriate.\(^{12}\)

\(^{12}\)We show in Appendix A that there is no ”reverted” signaling equilibrium in which unbiased advisers who actually prefer option \(A\) choose something different in O, just due to image concerns of being perceived as biased.
In the BONUS treatment, the second recommendation does not affect the adviser’s pecuniary payoff because the bonus has been removed. However, the previous recommendation might have been biased so that image concerns matter. An unbiased adviser should just recommend what he actually considers appropriate and, therefore, should repeat his initial advice. This means that, in order not to be perceived as biased, advisers who have previously been corrupted 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 appropriate. This means that these mimicking advisers re-recommend this option A in R2, even though they do not consider it appropriate advice. In addition, those advisers who truly prefer to recommend option A re-recommend this option in both, BONUS and NO BONUS. This then yields the following prediction:

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

Conditional on a scenario in which at least some advisers are corrupted by the bonus, thus that Prediction 1 is true, our design so far enables us to answer two main questions. By testing Prediction 3 we can detect persistent bias in advice-giving as predicted by our theory based on image concerns. If this prediction is confirmed, Prediction 2 enables us to investigate this effect in more detail. In case that Prediction 2a is correct, this would suggest that appropriate advice is determined based on advisers’ own preferences. When own choices and appropriate advice are not related, e.g. through self-serving beliefs about the clients’ preferences, we would expect Prediction 2b to be confirmed instead.

### 5.2 Predictions for the treatments BONUS vs. ANTICIPATE

The only difference between BONUS and ANTICIPATE is that the latter treatment gave advisers the possibility to anticipate the upcoming decisions they had to make before they issued their first recommendation. In essence, this allows advisers to consider ex-ante the decisions for R1, O, and R2 as one sequence of decisions with three interacting elements. Consequently, predictions in the ANTICIPATE-treatment depend on the assumptions regarding how advisers deal with backward-looking image concerns which can occur in the future. There are two principal possibilities which encompass the range of forward-looking behaviors. The first is that advisers factor in image concerns which arise from acting inconsistently into their decision-making process. The second is that, when they make the decision for R1, they disregard the such image concerns.

---

13 As for the own choice O we show in Appendix A that within our model, there is no “reverted” signaling equilibrium in which incorruptible advisers’ actions are affected by image concerns.
If advisers do ex-ante factor in these image costs, this puts an additional constraint on giving biased advice for option A in R1, due to one of two kinds of costs: One possibility is that they factor in the image costs when they plan to subsequently choose or re-recommend inconsistently in O or R2 as this signals their corruptibility. Alternatively, they can plan to act consistently, but this creates anticipated costs of having to choose sub-optimally for themselves and giving biased advice again. While there are these additional costs in ANTICIPATE, the reward for giving biased advice in R1 remains constant, relative to BONUS. This leads to the following prediction:

**Prediction 4a.** In R1, if advisers factor in image concerns ex-ante, they recommend option A more often in BONUS than in ANTICIPATE.

Note that, if advisers factor in image concerns ex-ante, the predicted consequences of the possibility to anticipate for O and R2 are not as clear as they are for R1. This is because of the following two countervailing forces: On the one hand, the previous prediction implies that fewer advisers recommend option A in R1. This decreases the share of those who might have to make persistently biased choices and recommendations in O and R2, in order to prevent signaling their corruptibility (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 acting inconsistently or choosing sub-optimally in O and 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). In consequence, it depends on the relative sizes of these two effects to determine the sign of the overall effect of how often option A is chosen and re-recommended in O and R2 if advisers factor in image concerns.

This is different if advisers do not factor in the image costs of giving inappropriate advice when they form a plan of actions ex-ante. For example, there is evidence that giving people the opportunity to consider their decisions ex-ante in a strategic manner can dampen pro-social behavior (for an overview, see Cooper and Kagel, 2016, pp. 230-231). Absent the considerations of upcoming image costs, the decision process underlying the first recommendation in ANTICIPATE then resembles the one in BONUS. It is therefore determined by a simple trade-off between the costs of giving inappropriate advice and earning the bonus. This implies similar rates of recommendations for option A in R1 between these treatments. However, it also has implications for recommendations in R2. While unanticipated backward-looking image concerns can create a pressure to act consistently in BONUS, negligence of these concerns when a plan of actions is formed ex-ante in ANTICIPATE means that this plan does not
feature persistent biases in R2. By analogous reasoning, the same applies for own choices in O if they have diagnostic value (i.e., if Prediction 2a is true). To the extent that advisers then follow through with this plan, the following prediction emerges:

**Prediction 4b. If advisers do not factor in image concerns ex-ante,**
- in R1, they recommend option A as often in ANTICIPATE as in BONUS,
- in O and if Prediction 2a is true, they choose option A less often in ANTICIPATE than in BONUS,
- in R2, they recommend option A less often in ANTICIPATE than in BONUS.

The above prediction is based on advisers who do not factor in image concerns, ex-ante form a plan (for R1, O, and R2), and 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 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, they 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 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 the two predictions is found to be true this also allows to rule out some alternative explanations, an issue which we consider in detail in Section 7. Thus, the comparison of ANTICIPATE and BONUS can bring further light on the determinants of biased advice-giving, in particular the role of backward-looking image-concerns therein.

### 5.3 Related empirical findings

Before we turn to our results, it is worth to note that recent empirical findings by others resonate with some of our predictions. Foerster et al. (2017)’s finding that adviser and not client characteristics largely determine financial advice supports the reasoning underlying Prediction 2a. The result by Linnainmaa et al. (2016) that the same advisers do actually choose the same portfolios they advise to clients for themselves – even after they went into retirement and even though these portfolios’ performance is, on average, worse than the market benchmark – then provides a real-world example which is in line with that prediction.

In a related lab experiment, Gneezy et al. (2016) report behavior which is also consistent with the above predictions, in particular those regarding the comparison of NO BONUS to BONUS. In fact, our experimental design for the first recommendation R1 in the BONUS-treatment is inspired by one of the
experiments they report. In its "before"-condition, advisers were first shown two investment options, call them option $X$ and $Y$. Before advisers got further information on the options, they were informed that they would get a bonus if they recommend option $X$ (which had a lower expected value but less variance than its alternative, option $Y$). Advisers were then explicitly asked to consider which one they would recommend to a client. Afterwards, advisers had to give an actual recommendation to a client.\footnote{They asked the advisers "Please take a minute to decide which product to recommend" (Gneezy et al., 2016, p.46) before advisers had to click a button stating that they were ready to issue their recommendation. Also note that in their experiment, these options were also called option $A$ and $B$, respectively, as in ours, but had different parameters. To avoid confusion, we call their options $X$ and $Y$.}

In their "after"-condition, advisers learned about the bonus after having initially considered which option to recommend but before actually recommending it. Gneezy et al. find that the bias towards option $X$ is larger in the "before" than in the "after"-condition. Their finding is consistent with the mechanism we propose. To see this, think of the initial consideration as an initial recommendation. Consider an adviser in their "after"-condition who has initially considered option $Y$ as appropriate and then learns about the bonus. If option $X$ is then recommended just to earn the bonus, this creates a clear (self-)signal pointing towards the adviser’s corruptibility, similar to a change in the first and second recommendation in our experiment. In contrast, advisers in the "before"-condition who know about the bonus and want to recommend option $X$ can adjust their initial consideration accordingly. This prevents signaling their corruptibility so that image costs of recommending option $X$ are lower than in the "after"-condition.\footnote{In a different study, Gneezy et al. (2018) expose referees in a joke-writing-contest to bribes by the contestants. In their "KeepWinner"-treatment, contestants submit their jokes together with bribes whereas in their "KeepWinnerDelayed"-treatment, workers submit bribes after their jokes were submitted and initially screened. They find a higher distortion in referee judgments in the former treatment, reflecting a similar effect as the above-described "before"/"after"-comparison.}

6 Results

6.1 Results for advisers’ 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 1 portrays the differences in 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 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 highly significant (Fisher exact test, BONUS vs. NO BONUS: \( p < 0.001 \), ANTICIPATE vs. NO BONUS: \( p < 0.001 \)). In contrast, the share of recommendations for option \( A \) in the two treatments which paid a bonus do not differ significantly (Fisher exact test, BONUS vs. ANTICIPATE: \( p = 0.843 \)).

We also employ a parametric approach by estimating the following regression model which includes additional control variables:

\[
I[r_{1,i} = A] = \alpha + \beta \cdot BONUS_i + \gamma \cdot ANTICIPATE_i + \delta \cdot c_i + \epsilon_i
\]  

(1)

In the above, the dependent variable is an indicator which 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 which indicate whether this subject was assigned to the respective treatments as opposed to NO BONUS, the baseline. The vector \( c_i \) collects control variables which indicate the subject's age, gender, monthly available budget, region of origin, the highest degree the subject holds or pursues and the field of studies.

Table 2 presents the results of this linear probability model, first without controlling for subject characteristics and then with such controls. Again, it is shown that the bonus leads to an increase of about 50 percentage points in the probability of recommending option \( A \) in the treatments BONUS and ANTICIPATE. Also reflecting the previous non-parametric results, this increase is not statistically different between these two treatments, as documented by the corresponding F-tests. The effect of the BONUS-treatment is in line with Prediction 1 while the effect of the bonus in ANTICIPATE supports Prediction 4b. The fact that the bonus affects recommendation equally strong in both treatments,
Table 2. Effect of bonus on advisors’ first recommendations in R1

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r_1 = \text{A} ) (first recommendation for option ( \text{A} ))</td>
<td>0.502***</td>
<td>0.448***</td>
</tr>
<tr>
<td>(0.078)</td>
<td>(0.086)</td>
<td></td>
</tr>
<tr>
<td>BONUS</td>
<td>0.481***</td>
<td>0.499***</td>
</tr>
<tr>
<td>(0.076)</td>
<td>(0.086)</td>
<td></td>
</tr>
<tr>
<td>Constant (NO BONUS)</td>
<td>0.039</td>
<td>-0.086</td>
</tr>
<tr>
<td>(0.027)</td>
<td>(0.228)</td>
<td></td>
</tr>
</tbody>
</table>

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 \( \text{A} \) in R1. The main independent variables are dummies indicating whether the adviser was in treatment B(=BONUS) or A(=ANTICIPATE) and whether option \( \text{A} \) was recommended in the first recommendation R1; N(=NO BONUS) is the reference category. Additional independent variables control for advisors’ 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 treatments BONUS and ANTICIPATE (45.8% and 48.0%, respectively) did not recommend option \( \text{A} \), even though they were offered money to do so. This is consistent with the notion that there exist non-pecuniary costs of giving such advice and that for a considerable fraction of advisers, these costs outweighed the pecuniary utility of the bonus.

6.2 Results for advisers’ own choices (O)

For their own choice, no bonus was paid to advisers in any condition. Figure 2 displays their choices for option \( \text{A} \) across treatments. In the baseline NO BONUS, we observe that 9.8% chose option \( \text{A} \) for themselves. This share is comparable to the results of Holt and Laury (2002) who find that six to eight percent of subjects exhibit risk-seeking preferences. In BONUS advisers were previously offered the bonus for their first recommendation. The share of those who chose option \( \text{A} \) then rises to 27.1%, almost three times as many advisers as in NO BONUS. This 17.3 percentage points increase is statistically
significant (Fisher exact test: \( p = 0.036 \)). The previously offered bonus does therefore continue to affect choices in this treatment where subsequent actions could not be anticipated. This is different in treatment ANTICIPATE. 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: \( p = 1.000 \)) and the findings in the previous literature.

Again, these findings are also mirrored in a regression analysis. For this, we replace the dependent variable in regression model (1) with a dummy indicating whether an adviser chooses option \( A \) for himself. Columns 1 and 2 in table 3 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 those who could not anticipate this stage but not for those who could. We therefore regard Prediction 2a as supported by our results for BONUS while the results for ANTICIPATE are in line with Prediction 4b.

Given these findings, it is helpful to recall the mechanism which underlies our reasoning concerning a persistent bias. It argues that if advisers base what they consider impartial advice on their own preferences, then they have to act according to their biased, previous advice in order to not signal the fact that they were corrupted. Therefore, the root cause for 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 subsequent own choice in O.

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 1) is included as an additional independent variable. If the bonus’ lasting effect on 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 3 shows that exactly this happens:
Table 3. Effect of bonus on advisors’ own choices in O

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>o = A (own choice for option A)</td>
<td>0.173**</td>
<td>0.181**</td>
<td>0.024</td>
<td>0.351***</td>
<td>0.384**</td>
</tr>
<tr>
<td>(0.077)</td>
<td>(0.084)</td>
<td>(0.076)</td>
<td>(0.078)</td>
<td>(0.150)</td>
<td>(0.145)</td>
</tr>
<tr>
<td>BONUS</td>
<td>0.018</td>
<td>0.009</td>
<td>-0.184***</td>
<td>-0.119</td>
<td></td>
</tr>
<tr>
<td>(0.057)</td>
<td>(0.066)</td>
<td>(0.067)</td>
<td>(0.145)</td>
<td>(0.145)</td>
<td>(0.145)</td>
</tr>
<tr>
<td>ANTICIPATE</td>
<td>0.098**</td>
<td>-0.018</td>
<td>0.012</td>
<td>-0.390</td>
<td>0.177</td>
</tr>
<tr>
<td>(0.042)</td>
<td>(0.175)</td>
<td>(0.149)</td>
<td>(0.316)</td>
<td>(0.168)</td>
<td>(0.168)</td>
</tr>
<tr>
<td>Constant (NO BONUS)</td>
<td>6.390**</td>
<td>5.600**</td>
<td>7.360***</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(no)</td>
<td>(yes)</td>
<td>(yes)</td>
<td>(yes)</td>
<td>(yes)</td>
<td>(yes)</td>
</tr>
<tr>
<td>Controls</td>
<td>OLS</td>
<td>OLS</td>
<td>OLS</td>
<td>2SLS</td>
<td>2SLS</td>
</tr>
<tr>
<td>Data used from treatments</td>
<td>B,A,N</td>
<td>B,A,N</td>
<td>B,A,N</td>
<td>B,N</td>
<td>A,N</td>
</tr>
<tr>
<td>Observations</td>
<td>149</td>
<td>149</td>
<td>149</td>
<td>99</td>
<td>101</td>
</tr>
</tbody>
</table>

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).

The previously positive and statistically significant coefficient for BONUS becomes essentially zero and insignificant while the coefficient on \( r_1 = A \) takes up all the explanatory power. Also note that even if one controls for the initial recommendation for option A, the implied rate of own choices in BONUS is still significantly larger than in ANTICIPATE.

We can also quantify more exactly the mediating effect which bonus-induced first recommendations for option A had on advisers’ own choices 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 own choices in BONUS by its effect on the initial recommendations. This assumes that this channel is the only way through which the bonus affected subsequent own choices. From
the unconditional estimates in the first columns of Table 2 and 3 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 before-hand. We then get from these numbers that 34.4% (\(\approx 0.173/0.502\)) of the advisers whose initial advice was shifted towards option A in BONUS also adjusted their own choices accordingly.

Note that the above estimate is equivalent to the Wald-estimate 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 recommendations for option A in R1. Then, based on this first stage, the bonus-induced effect of the initial recommendation on own choices is estimated in the second stage. In the fourth column of Table 3. we present these second stage results, i.e., the local average treatment effect-estimates when additional controls are added. When we add these controls, the mediating effect of the bonus’ influence on initial recommendations, modeled explicitly through the initial presence of the bonus, increases to a significant shift of 38.4 percentage points in the probability of later choosing the risky option for oneself.

When the same method is used to investigate whether the initial bonus also affected advisers’ subsequent own choices when they could anticipate the upcoming choices they had to make, the results look different. Even though the first-stage result of the effect of the bonus on first recommendations are similar between BONUS and ANTICIPATE, they do not spill over on own choices in the later case. Column 5 of Table 3 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 subsequent own choices is comparatively small in magnitude and not statistically different from zero. Again, these results are in line with Prediction 4b. This suggests that advisers do not factor in the image aspect of their choices when they evaluate the sequence of actions ex-ante.

6.3 Results for advisers’ 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 3 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 6 below).

This is very different for second recommendations in BONUS. Although there is no bonus in R2 anymore, 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, a significant increase by 19.0 percentage points relative to NO BONUS (Fisher exact test: $p = 0.007$). 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 and the choice they had to make for themselves. 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 (1), now with a dummy which indicates whether option A is recommended in the second recommendation as the dependent variable. Columns 1 and 2 in Table 4 present the results and show that the effect of the BONUS-treatment remains largely unchanged, independently of whether controls are added. The finding that this happens when the adviser’s future actions and the bonus’ removal were unexpected but not when they could be anticipated is also mirrored in these regression results which control for remaining heterogeneity. These results therefore support Prediction 3 and Prediction 4b.

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 4 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 which 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 4 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 5 then shows that, similar as for own choices, this spill-over of giving in to the bonus in R1 does not
<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_2 = A$ (second recommendation for option $A$)</td>
<td>$0.190^{***}$</td>
<td>$0.193^{**}$</td>
<td>0.092</td>
<td>0.039</td>
<td>-0.199</td>
</tr>
<tr>
<td></td>
<td>$(0.067)$</td>
<td>$(0.078)$</td>
<td>(0.080)</td>
<td>(0.027)</td>
<td>(0.176)</td>
</tr>
<tr>
<td>BONUS</td>
<td>0.021</td>
<td>0.055</td>
<td>-0.058</td>
<td>0.039</td>
<td>-0.199</td>
</tr>
<tr>
<td></td>
<td>$(0.044)$</td>
<td>$(0.055)$</td>
<td>(0.072)</td>
<td>(0.027)</td>
<td>(0.176)</td>
</tr>
<tr>
<td>$r_1 = A$</td>
<td>0.227***</td>
<td>0.227***</td>
<td>0.415***</td>
<td>0.039</td>
<td>-0.199</td>
</tr>
<tr>
<td></td>
<td>$(0.080)$</td>
<td>$(0.080)$</td>
<td>(0.147)</td>
<td>(0.027)</td>
<td>(0.176)</td>
</tr>
<tr>
<td>$\hat{r}_1 = A$ (via BONUS)</td>
<td>0.415***</td>
<td>0.415***</td>
<td>0.008</td>
<td>0.039</td>
<td>-0.199</td>
</tr>
<tr>
<td></td>
<td>$(0.147)$</td>
<td>$(0.147)$</td>
<td>(0.101)</td>
<td>(0.027)</td>
<td>(0.176)</td>
</tr>
<tr>
<td>$\hat{r}_1 = A$ (via ANTICIPATE)</td>
<td>0.008</td>
<td>0.008</td>
<td>0.072</td>
<td>0.039</td>
<td>-0.199</td>
</tr>
<tr>
<td></td>
<td>$(0.101)$</td>
<td>$(0.101)$</td>
<td>(0.146)</td>
<td>(0.027)</td>
<td>(0.176)</td>
</tr>
<tr>
<td>Constant (NO BONUS)</td>
<td>0.039</td>
<td>-0.199</td>
<td>-0.180</td>
<td>-0.331</td>
<td>-0.072</td>
</tr>
<tr>
<td></td>
<td>$(0.027)$</td>
<td>$(0.176)$</td>
<td>(0.181)</td>
<td>(0.242)</td>
<td>(0.146)</td>
</tr>
<tr>
<td>F-test: BONUS=ANTICIPATE</td>
<td>5.830**</td>
<td>3.430*</td>
<td>3.960**</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Controls</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Estimation method</td>
<td>OLS</td>
<td>OLS</td>
<td>OLS</td>
<td>2SLS</td>
<td>2SLS</td>
</tr>
<tr>
<td>Data used from treatments</td>
<td>B,A,N</td>
<td>B,A,N</td>
<td>B,A,N</td>
<td>B,N</td>
<td>A,N</td>
</tr>
<tr>
<td>Observations</td>
<td>149</td>
<td>149</td>
<td>149</td>
<td>99</td>
<td>101</td>
</tr>
</tbody>
</table>

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).

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 4b.

### 6.4 Further results

There are some additional findings which 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. Table 5 shows the frequency of advisers’ own choices over their first recommendations. For NO BONUS, when there is no incentive to be inconsistent, 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 in line 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 5 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 4b 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 size in BONUS.

The consistency-pattern between advisers’ first and second recommendations, displayed in Table 6, 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 6’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 either on the diagonal or the first row of Table 6’s middle three columns. Again, the picture is similar for treatment ANTICIPATE, where a total of 10.0% of the 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.

Table 5. Frequencies of advisers’ own choices conditional on their first recommendation

| R1 | O   | NO BONUS |  | BONUS |  | ANTICIPATE |
|----|-----|----------|  |       |  |           |
|    | A   | 3.9%     |  | 22.9% |  | 8.0%      |
|    | B   | 2.0%     |  | 0.0%  |  | 0.0%      |
|    | C   | 3.9%     |  | 4.2%  |  | 0.0%      |

26
Table 6. Frequencies of advisers’ second recommendations conditional on their first recommendation

<table>
<thead>
<tr>
<th></th>
<th>NO BONUS</th>
<th></th>
<th>BONUS</th>
<th></th>
<th>ANTI-PARE</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>A</td>
<td>3.9%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>18.7%</td>
<td>16.7%</td>
<td>18.8%</td>
</tr>
<tr>
<td>B</td>
<td>0.0%</td>
<td>35.3%</td>
<td>2.0%</td>
<td>0.0%</td>
<td>6.2%</td>
<td>0.0%</td>
</tr>
<tr>
<td>C</td>
<td>0.0%</td>
<td>15.7%</td>
<td>43.1%</td>
<td>4.2%</td>
<td>8.3%</td>
<td>27.1%</td>
</tr>
</tbody>
</table>

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 ANTI-PARE, 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: $p = 0.533$).

These results become stronger, both numerically and statistically, if they are regarded in a regression framework. Columns 1 and 2 of Table 7 present the results from estimating model (1) 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 ANTI-PARE.

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 control 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

---

17Due 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 session. When we exclude these sessions 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>
<thead>
<tr>
<th>Dependent variable</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r_1 = A ) (via BONUS)</td>
<td>0.914**</td>
<td>1.066**</td>
<td>0.343</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( r_1 = A ) (via ANTICIPATE)</td>
<td>0.299</td>
<td>0.151</td>
<td>-0.655</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant (NO BONUS)</td>
<td>4.960***</td>
<td>6.155***</td>
<td>5.822***</td>
<td>7.635***</td>
<td>5.065***</td>
</tr>
<tr>
<td>F-test: BONUS=ANTICIPATE</td>
<td>1.860</td>
<td>3.900**</td>
<td>5.450**</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Controls</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Estimation method</td>
<td>OLS</td>
<td>OLS</td>
<td>OLS</td>
<td>2SLS</td>
<td>2SLS</td>
</tr>
<tr>
<td>Data used from treatments</td>
<td>B,A,N</td>
<td>B,A,N</td>
<td>B,A,N</td>
<td>B,N</td>
<td>A,N</td>
</tr>
<tr>
<td>Observations</td>
<td>149</td>
<td>149</td>
<td>149</td>
<td>99</td>
<td>101</td>
</tr>
</tbody>
</table>

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).

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. In fact, the pattern observed for the answers to this general risk-question mirrors exactly the one observed for own choices and repeated recommendations.
7 Discussion

Our results show that incentives to bias advice can have a lasting and causal effect on both, advisers’ future recommendations and their own choices for risky decisions. This behavior is consistent with a psychological mechanism we propose and which captures the insight that changing advice after a conflict of interest disappeared signals one’s corruptibility. This mechanism also assumes that issuing biased advice is costly. Only when these costs are low enough, relative to image costs, advisers stick to their initial, biased recommendation. In line with this, our estimates imply that around 40% of those advisers whose initial advice was biased by the bonus changed their recommendation when they had to give it again after the bonus was removed.

We find similar effects with regard to advisers who chose the risky option for themselves because the bonus induced them to initially recommend it. This allows to differentiate how advisers determine appropriate advice and what signals unbiased advice. If advisers had formed a motivated belief about their clients’ preferences independent of their own, this would not have required them to choose consistently for themselves to avoid signaling their corruptibility. However, if what is considered appropriate advice is based on own preferences, it is necessary to actually choose as one has previously advised in order to avoid such a signal. Our results suggest that the latter reasoning is relevant.

When the bonus is removed but advisers knew ex-ante about this removal and the requirement to act after their initial recommendation, we do not observe a persistent effect. Although initial recommendations in R1 are as biased in ANTICIPATE as in BONUS, relative to when no bonus is present, its persistent effect on subsequent decisions manifests only in the latter treatment. This finding that the one-off bonus leads to persistent bias when its removal is unanticipated but not when advisers can foresee is consistent with our proposed theory. It suggests that when advisers form a plan ex-ante and follow through with it, they disregard backward-looking image costs whereas these matter when they have a less strategic, more retrospective perspective.

Our findings also allow to rule out alternative explanations for the persistent biases we find. For example, an explanation in which advisers are clueless about which option to take and then just take the bonus as a cue for what is the ”best” option makes different predictions. First, all advisers whose initial advice was affected by the bonus should then stick to their cue-induced initial recommendation when they make a second recommendation or a choice for themselves. The partial consistency (about 40%) we find speaks against such a mechanism. Further, direct evidence against such a ”cue”-effect
comes from the behavioral differences we observe between BONUS and ANTICIPATE. This is because if advisers take the bonus as a cue, it should have the same influence on later decisions, independently of whether they were anticipated or not. This applies especially because we find that the bonus’ effect on the first recommendation is virtually the same in both of these treatments – a persistent, cue-induced effect in the decisions which followed should therefore be equally strong. However, the observation that partial consistency occurred in BONUS but not in ANTICIPATE contradicts the notion that subjects just follow mechanically the option to which the bonus pointed. Related alternative explanations based on the notion that the bonus leads to initial recommendations for option A which then serve as an anchor (Tversky and Kahneman, 1974) or are understood as an experimenter demand (Zizzo, 2010; de Quidt et al., 2017) towards this option to which subjects then (partially) stick can be refuted by the same reasoning.\footnote{The results also rule out decreasing absolute risk aversion as another potential explanation. Such an explanation would argue that advisers become more risk-seeking after earning the 3 GBP. Again, this should then apply independently of whether they are in BONUS or ANTICIPATE.}

Additional evidence comes from subjects’ self-stated willingness to take risks. Their responses mirror the results for the bonus’ effect on own choices and repeated recommendations. In particular, having recommended the risky option for a bonus led advisers to state a higher general preference for risk when they could not anticipate the upcoming decisions they had to make. Again, this shows that the subset of biased advisers who were persistently biased did not act mechanically when they chose for themselves or re-recommended from the same set of possible options. Rather, this result shows that their initial, biased recommendations have signaling implications which apply to 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. Such concerns could be driven by advisers trying to maintain a positive self-image as in dual-self models. Such models capture the notion that one constantly learns through one’s actions about the own underlying motives by modeling this inferring self as an outside observer. Given our experimental setup, we interpret these findings as in line with such self-image concerns.\footnote{In our experiment, social image concerns were limited by the fact that advisers wrote down their recommendations and choices in private and put them in envelopes. Thus, even the experimenter could not be expected to see the sequence of advisers’ actions and choices during the experiment.} However, social image concerns with regard to an actual outside observer follow the same mechanics and could therefore lead to the same persistent effects as long as this observer sees the sequence of an adviser’s choices and recommendations (e.g. supervisors or regulators overseeing adviser behavior after a new law bans commission-based advice). The findings that the bonus’ persistent effect
depends on the possibility for advisers to anticipate its one-off nature, has further important implications for the governance of adviser-based industries which we discuss below.

8 Conclusion

Our findings have several immediate implications. First, we present evidence that biases in advice-giving can loom longer than the conflict of interest which caused them. Recent policies which ban the causes of conflicts of interest are certainly a right 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 those who have been exposed to it.

Second, we also find these persistent effects on advisers’ recommendations after incentives to bias them were removed, even though they had to choose for themselves before. This observation speaks against the general potential of just letting advisers 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 those who advise external clients when the same persons, for example financial analysts, also affect related decisions within the company. In addition, we propose a mechanism which is capable to explain these results and those observed by others in both, lab experiments and field data.

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 been known to the adviser from the beginning. While this does not diminish the bias on initial advice, 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: While 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 my be different for more experienced ones. Experienced advisers might have spent a considerable part of their professional career being exposed to such incentives. Their removal, even when announced before-hand, may thus come comparatively surprising for them. In addition, image concerns loom larger for them as they threaten a considerable part of such advisers’ professional identities. Also, our results suggest that while the possibility to anticipate the transitory nature of a conflict of interest can prevent a persistent bias in future actions, it does not diminish bias in current advice when the conflict of interest is still present.
Finally, it is important to note that while our findings are on advice for risky choices, they are not necessarily bounded to this specific domain. The crucial feature is that there is no clear-cut right or wrong so that one can reasonably maintain the image that advice was genuine and unbiased, even though it was not. 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


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 signaling model of persistent biases in advice-giving

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 $C$ 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 $B \subset C$. Hence, when $b > 0$, the adviser is subject to a bias towards recommending a choice from $B$.$^1$ We denote by $N$ the subset of choices with no bonus, i.e., $N = C/B$. Both, $B$ and $N$ 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 $u : \mathbb{R} \to \mathbb{R}$ with $u(0) = 0$. We also assume that all choices in $C$ can be represented by a pecuniary payoff. This payoff is captured by the function $v : C \to \mathbb{R}$, for example via the corresponding certainty equivalent when elements of $C$ are options. We omit the $v$-function if it is the identity function. The own choice $o$ which an adviser optimally chooses for himself from a, possibly restricted, subset $X \subseteq C$ is denoted by $o^*_X \equiv \max_{o \in X} \{v(o)\}$. To save on notation we assume w.l.o.g. that $o^*_X$ is a singleton for each $X \subseteq C$. The subscript is omitted when the choice-set is unrestricted, i.e., $o^*_X = o^*_C$. The share of advisers for whom $o^*_X \in X \subseteq C$, i.e., whose unconstrained optimum lies in $X$ will be denoted with $\alpha_X$ (thus, for such advisers $o^*_X = o^*_X$ holds).

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

- **Projected appropriate recommendations:** 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 mean that $r^* = o^*$ holds and therefore $\alpha_X = \beta_X$ for each $X \subseteq C$.

- **Independent appropriate recommendations:** 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, such a belief

$^1$Note that this is equivalent to a punishment $p = -b$ he has to pay if he does not recommend an option from $B$. 


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 \( r \in C \) when \( r \neq r^* \).

**Image costs of being perceived as biased:** In addition to the immediate costs 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:** Let \( \pi(a) \) denote the corresponding pecuniary payoff which an adviser gets from action \( a \in C \) which is either a recommendation or a choice for himself. Specifically,

\[
\pi(a) = \begin{cases} 
   v(a) & \text{if } a = o \text{ is the adviser’s own choice for himself;} \\
   b \cdot 1[a \in B] & \text{if } a = r \text{ is the adviser’s recommendation to a client.}
\end{cases}
\]

where \( 1[\cdot] \) denotes the indicator function which takes a value of one if the statement in the bracket is true. Given an adviser’s history \( h \) of previous own choices and recommendations plus the choice \( r^* \) he considers appropriate, his overall utility can then be written as follows:

\[
U(a \mid h, r^*) = u(\pi(a)) - \kappa \cdot 1[a \neq r^* \text{ and } a \text{ is a recommendation}] - \lambda \cdot Pr[\text{previous advice was biased} \mid a, h]
\]

In the above, 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. Note that if \( Pr[\text{previous advice was biased} \mid a, h] \) varies with \( a \) (“\( a \) has diagnostic value”), choosing an action \( a \) which maximizes (1) corresponds to playing a signaling game. To
solve such a game, we use Perfect Bayesian Equilibrium as a solution concept: Given a prior history \( h \), the adviser chooses \( a \) such that \( U(a \mid h, r^*) \) is maximized given that \( \Pr[\text{previous advice was biased} \mid a, h] \) is updated via Bayes’ rule under knowledge of the adviser’s strategy. 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 \( \kappa \) and \( \lambda \) are independent of \( o^* \) and \( r^* \).

\(^2\)With such correlation all our results would remain valid if the joint distributions of \( (\kappa, \lambda) \), 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 \( c \in C \) would be such a sufficient (but not necessary) condition.

**Lemma 1.** Suppose the joint distribution of \( (\kappa, \lambda) \) is absolutely continuous and the associated p.d.f. has full support over \( \mathbb{R}_+^2 \times \mathbb{R}_+^2 \). 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. \( \Lambda(y \mid x) = \Pr[\lambda \leq y \mid \kappa \leq x] \) 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 \( \{\alpha_X\}_{X \subseteq C} \) and \( \{\beta_X\}_{X \subseteq C} \) 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., \( \beta^C \in (0, 1) \). While an adviser knows his individual values of \( (\kappa, \lambda, r^*) \), 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.

**A2. Analysis of BONUS and NO BONUS**

We will now analyze how a one-off incentive can lead to a persistent bias in advice-giving. For this, we consider a situation in which an adviser first has to issue a recommendation \( r_1 \in C \) for which he can earn a bonus \( b \) and then a second recommendation \( r_2 \in C \) to another client for which no bonus can be earned.

Our experiment resembles this setting with \( C = \{A, B, C\} \) and \( B = \{A\} \) 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 \( o \in C \) 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 \( \pi(r_1) = 0 \) because \( b = 0 \), (1) becomes \( U(r_1 \mid r^*) = -\kappa \cdot \mathbb{1}[r_1 \neq r^*] \). Accordingly, advisers recommend \( r_1 = r^* \) and the share of advisers who recommend an option from \( B \) is given by \( \beta_B \).

R1 – BONUS: Recommending an option from \( B \) now yields the bonus, captured by \( \pi(r_1) = b \cdot \mathbb{1}[r_1 \in B] \) with \( b > 0 \). There is no previous advice, so that image concerns do not matter. Thus, (1) becomes \( U(r_1 \mid r^*) = u(b \cdot \mathbb{1}[r_1 \in B]) - \kappa \cdot \mathbb{1}[r_1 \neq r^*] \). For the share \( \beta_B \) of advisers who have \( r^* \in B \), recommending \( r_1 = r^* \) is then clearly optimal – they get rewarded for what they would have recommended anyway. However, for a share \( 1 - \beta_B \) of advisers, \( r^* \in \mathcal{N} \), holds. They face a trade-off between recommending an option from \( B \) even though they do not consider it appropriate and being impartial. Recommending an option from \( B \) yields them pecuniary utility \( u(b) \) but causes costs \( \kappa \) 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 \( \kappa \) lower than \( u(b) \) give biased advice; their population share is given by \( (1 - \beta_B) \cdot K(u(b)) > 0 \). 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 \( \kappa, \lambda, \) and \( r^* \):

- \( \theta = 1 \) – incorruptible advisers who recommend an option which earns them a bonus because they truly think that it is appropriate for the client \( (r_1 = r^* \in B) \). Their population share is \( \phi_1 \equiv \beta_B \).

- \( \theta = 2 \) – incorruptible 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 corrupted by the bonus because their \( \kappa \) is sufficiently high. Their population share is \( \phi_2 \equiv (1 - \beta_B) \cdot (1 - K(u(b))) > 0 \).

4
• θ = 3 – corruptible advisers who recommend an option which earns them a bonus even though they do not think that it is appropriate to do so (r1 ∈ B but r1 ≠ r* ∈ N). They are corrupted by the bonus because their κ is low enough. Their population share is φ3 ≡ (1 − βB) · K(u(b)) > 0.

Note that by letting b = 0, the above also applies to the NO BONUS-treatment. In this case, only share φ1 recommends an option from B as there are no type-3-advisers. Also note that from the above, Pr[previous advice was biased | a, h] = Pr[θ = 3 | a, h] 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 B in BONUS is given by φ1 + φ3 and is larger than the share φ1 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 can be expressed by π(o) = v(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 | r1, r*) = u(v(o)) which is maximized by an adviser’s preferred own choice o*. Thus, the share of advisers choosing an option from B is given by αB.

O – BONUS: As there were corruptible advisers in the previous recommendation R1 (the type-3-advisers) image costs of being perceived as them matter. Given the prior history h = r1 and the current choice for oneself a = o, the adviser’s objective function (1) becomes U(o | r1, o, r*) = u(v(o)) − λ · Pr[θ = 3 | r1, 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 r1 ≠ o* = r* would be put on the spot: By choosing o = o* ≠ r1 they would reveal themselves as type-3-advisers with r1 ≠ r* because type-1 and type-2-advisers choose o = o* = r1 = 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 = r1 ∈ B. By this, they would lower the weight on the image costs of being perceived as corruptible but incur costs of not choosing what they actually prefer because for them, o* ∈ 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 θ ∈ {1, 2} and a share π0* ∈ (0, 1) of advisers of type θ = 3 choose o = r1. The remaining share of type-3-advisers chooses o ≠ r1.

**Proof:** see Appendix B.
Independent appropriate recommendations: Having a (possibly self-serving) belief about what constitutes appavoid 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., $Pr[\theta = 3 \mid r_1, o]$ is invariant to $o$. The choice $o \in C$ which maximizes $U(o \mid r_1, o, r^*) = u(v(o)) - \lambda \cdot Pr[\theta = 3 \mid r_1, o]$ 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 $o \in B$ in BONUS is given by $\phi_1 + \pi_3 \phi_3$ with $\pi_3 \in (0, 1)$ and is larger than $\phi_1$, 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 $o \in B$ is given by $\phi_1$ 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 $\pi(r_2) = 0$. 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 $U(r_2 \mid r_1, r_2, r^*) = -\kappa \cdot 1_{[r_2 \neq r^*]}$. Recommending $r_2 = r^*$ maximizes this expression so that share $\phi_1$ of advisers recommend an option from $B$.

R2 – BONUS: In this condition, the initial bonus has been removed so that $\pi(r_2) = 0$ 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 $U(r_2 \mid r_1, r_2, r^*) = -\kappa \cdot 1_{[r_2 \neq r^*]} - \lambda \cdot Pr[\theta = 3 \mid r_2, r_1, o]$.

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 $r_2 = r_1 \in B$ in order to pool with type-1-advisers which discounts their image costs $\lambda$. However, since for them $r^* \in N$ holds, they would suffer costs $\kappa$ of issuing biased advice (again). If they want to prevent these costs and recommit $r_2 = r^*$, this means that $r_2 \neq r_1$ so that they would reveal themselves as type-3-advisers and suffer full image costs $\lambda$. They do so if $\lambda$ is small, relative to the costs $\kappa$ 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 \( \pi^*_o \) 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 \( \psi \cdot \pi^*_\theta \) 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 \( \pi^*_\theta \in (0, 1] \). 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 \( B \) in BONUS is given by \( \phi_1 + \psi \pi^*_3 \phi_3 \) and is larger than \( \phi_1 \), 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 \((r_1, o, r_2) \in \{B, N\} \times \{B, N\} \times \{B, N\}\). 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 \((r_1, o, r_2) \in B \times N \times B\) is written simply as \( BNB \) and means that the first and second recommendation \( r_1 \) and \( r_2 \) are for an option from \( B \) whereas the own choice \( o \) is from \( N \).

When advisers form a plan for \( R_1, O, \) and \( R_2 \), it is unclear whether they account for backward-looking image implications of their actions in \( O \) and \( R_2 \), or whether their ex-ante, more strategic perspective leads them to ignore such concerns. Therefore, the predictions for the ANTICIPATE treatment are derived for two cases: (1) advisers who factor in the image implications of their decisions when they ex-ante form a plan; (2) advisers who do not factor in image implications of their (planned) decisions.

**Advisers factor in image concerns**

Suppose advisers do consider image concerns when they make a plan for their decisions. 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 together, as given by the following expression. In it, each line corresponds the effect on the payoff in \( R_1, O, \) and \( R_2 \), respectively:

\[
U((r_1, o, r_2) \mid h, r^*) = u(\pi(r_1)) - \kappa \cdot \mathbb{1}[r_1 \neq r^*] + u(o) - \lambda \cdot \Pr[\theta = 3 \mid r_1, o] - \kappa \cdot \mathbb{1}[r_2 \neq r^*] - \lambda \cdot \Pr[\theta = 3 \mid r_1, o, r_2]
\]
For ease of notation, we define the following abbreviations for the posterior that an adviser is type-3 (\( \theta = 3 \)) conditional on his actions \((r_1, o)\) or \((r_1, o, r_2)\), respectively:

\[
P_{M_1, M_2} \equiv \Pr[\theta = 3 \mid r_1 \in M_1, o \in M_2]
\]

\[
P_{M_1, M_2, M_3} \equiv \Pr[\theta = 3 \mid r_1 \in M_1, o \in M_2, r_2 \in M_3]
\]

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

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 \(o^*\) for the optimal choice from a (possibly restricted) subset \(X \subseteq C\). As before, we omit the subscript when the choice set is unrestricted, thus \(o^* = o^*_C\). 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 6.2) and because it considerably simplifies the exposition by limiting it to relevant cases.³

The payoffs of all possible sequences of actions for advisers with \(r^* \in B\) are as follows:

\[
BBB : \quad u(b) + u(o^*_B) - \lambda \cdot P_{BB} - \lambda \cdot P_{BBB}
\]

\[
BBN : \quad u(b) + u(o^*_B) - \lambda \cdot P_{BN} - \lambda \cdot P_{BBN}
\]

\[
BNN : \quad u(b) + u(o^*_N) - \lambda \cdot P_{BN} - \lambda \cdot P_{BNN}
\]

\[
BNB : \quad u(b) + u(o^*_N) - \lambda \cdot P_{BN} - \lambda \cdot P_{BNB}
\]

\[
NBN : \quad u(o^*_N) - 2\kappa
\]

\[
NB : \quad u(o^*_N) - \kappa
\]

\[
NBB : \quad u(o^*_B) - \kappa
\]

³We also have modeled the case when own choices and appropriate action are independent. In such a setting, the share of advisers who choose \(o \in B\) 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 \(r_1 \in B\) in \(R1\). In consequence, only the exact, numerical prediction but none of the following qualitative predictions for ANTICIPATE change (formal results are available upon request).
Comparing these payoffs then leads to Proposition 3. It shows that the share of advisers with $r^* \in \mathcal{N}$ who choose $r_1 \in B$ 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 $r^* \in B$ do not plan to recommend $r_1 \in B$ when they can anticipate upcoming choices. The intuition behind this result can be derived in four steps.

First, in a plan which features $r_1 \in B$ and $o \in B$, an adviser with $r^* \in B$ would always prefer to recommend $r_2 \in B$ 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 $BBB$ or $BBN$, depending on the relative size of their $\kappa$ and $\lambda$. In any case, $P_{BBN} = 1$ applies.

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

Third, given that $P_{BN} = 1$, advisers with $r^* \in B$ prefer $BBB$ over $BNB$ 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 $BNB$ as it is dominated by $BNN$. This is because a choice $o \in \mathcal{N}$ following a recommendation $r_1 \in B$ also reveals the adviser to be of type-3 while recommending $r_2 \in B$ does not elevate image concerns. Taken together, this means that advisers with $r^* \in B$ prefer $BBB$ over the rest of plans which feature $r_1 \in B$ while advisers with $r^* \in \mathcal{N}$ may prefer $BBB$, $BBN$, or $BNN$ over other plans which feature $r_1 \in B$.
Finally, after recommending $r_1 \in \mathcal{N}$, advisers with $r^* \in \mathcal{B}$ prefer $\mathcal{N}BB$ over other plans which feature $r_1 \in \mathcal{N}$. Alternatively, advisers with $r^* \in \mathcal{N}$ prefer $\mathcal{N}NN$ over the rest of plans which feature $r_1 \in \mathcal{N}$. These decisions are made because recommending $r_1 \in \mathcal{N}$ 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 $BBB$ or $NBB$ 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: $BBB$, $BBN$, $BNN$, and $NNN$. 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 3:

**Proposition 3.** Suppose advisers factor in image-concerns ex-ante. 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 $\beta_B$ – 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 $(1 - \beta_B)K(u(b))$, 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 4a in the main text:

**Corollary 4a.** If advisers factor in image costs, the share of them who recommend $r_1 \in \mathcal{B}$ in the ANTICIPATE treatment, given by $\beta_B \cdot \tilde{\tau}_{r_1}^* + (1 - \beta_B) \cdot \tilde{\pi}_{r_1}^*$, is strictly lower than the share of them in BONUS, given by $\beta_B + (1 - \beta_B)K(u(b))$.

**Advisers do not factor in image-concerns**

If advisers do not account for backward-looking image cost when they pick a sequence of actions ex-ante, the utility function of an adviser is given by

$$U((r_1, o, r_2) | h, r^*) = u(\pi(r_1)) - \kappa \cdot 1_{[r_1 \neq r^*]} + u(o) - \kappa \cdot 1_{[r_2 \neq r^*]}.$$
We start by considering the decision of advisers with \( r^* \in B \). It is straightforward to check that \( BBB \) dominates all alternatives since the corresponding payoff is \( u(b) + u(o^*) \) whereas all other choices involve giving inappropriate advice or making a suboptimal choice.

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

\[
\begin{align*}
BBB &: u(b) + u(o^*_B) - 2\kappa \\
BBN &: u(b) + u(o^*_B) - \kappa \\
BNB &: u(b) + u(o^*_B) - 2\kappa \\
BNN &: u(b) + u(o^*_B) - \kappa \\
N\!N\!N &: u(o^*_B) \\
N\!B\!N &: u(o^*_B) - \kappa \\
N\!B\!B &: u(o^*_B) - \kappa
\end{align*}
\]

Simple comparisons show that (21) dominates (18), (19), and (20) while (22) dominates (23), (24), and (25). Hence, these advisers would either choose \( BNN \) or \( NNN \). They choose the former over the latter whenever \( u(b) + u(o^*_B) - \kappa > u(o^*_B) \), leading to the following result:

**Proposition 4.** Suppose advisers do not factor in image-concerns. All advisers with \( r^* \in B \) choose \( BBB \). The share of advisers with \( r^* \in N \) choose \( BNN \) is given by \( K(u(b)) \). All other advisers with \( r^* \in N \) choose \( NNN \).

**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 \( r_1 \in B \). Also, a share \( K(u(b)) \) of advisers with \( r^* \in N \) recommend such an option. For the case of no bonus, we got that only type-1-advisers with \( r^* \in B \) choose \( o \in B \) and recommend \( r_2 \in B \).

The above proposition shows that when anticipation was possible, the behavior for type-1-advisers is the same. This implies the following, leading to Prediction 4b:

**Corollary 4b.** If advisers do not factor in image costs,

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

b) the share of advisers who choose \( o \in B \) in \( \text{ANTICIPATE} \) is given by \( \beta_B \) and equals the corresponding share in \( \text{NO BONUS} \) (and is therefore lower than in \( \text{BONUS} \)),

c) the share of advisers who recommend \( r_2 \in B \) in \( \text{ANTICIPATE} \) is given by \( \beta_B \) and equals the corresponding share in \( \text{NO BONUS} \) (and is therefore lower than in \( \text{BONUS} \)).
A4. Discussion of the model

The above analysis of behavior in BONUS relative to NO BONUS shows how, through image concerns of being perceived as biased, a one-off bonus can lead advisers to repeat biased advice. It can even lead them to choose for themselves in a way which, absent such concerns, would be sub-optimal. For this, an adviser has to be initially corrupted by a previous bonus, thus he has to have small enough costs $\kappa$ of giving inappropriate advice. The persistent effect then occurs when, in addition to sufficiently low values of $\kappa$, image costs, as measured by $\lambda$, are high enough.

The grey rectangle in Figure A.1 a) depicts the relevant parameter constellations such that own choices, when advisers cannot anticipate, are persistently affected by the bonus: Low enough costs to bias advice ($\kappa$ below the vertical line) and high enough image concerns ($\lambda$ above the horizontal line). Figure A.1 b) shows the parameter-values which lead to persistent bias in repeated advice. For this, advisers have to have sufficiently high image concerns such that they rather re-recommend an inappropriate choice than to change their advice and thereby reveal themselves as corruptible. Accordingly, image costs $\lambda$ have to be high relative to the costs $\kappa$ of issuing inappropriate advice, i.e., above the bold diagonal line.

When own choices have diagnostic value (projected appropriate recommendations), this is only relevant to those in the left grey rectangle who have not yet revealed themselves, i.e. those in the grey rectangle. Therefore, the relevant parameters for repeated biased advice when own choices have diagnostic value are within the dark grey pentagon. If own choices do not have diagnostic value (independent appropriate recommendations) the parameter restrictions on $\lambda$ in panel a) can be ignored and it is only panel b)’s diagonal (plus the vertical limit on $\kappa$ which ensures that biased advice was given initially) which sets the threshold for repeated biased advice. Note that the latter reasoning would also apply if there were only repeated advice-giving and no own choice.

In the model, image costs are defined rather broadly. They are measured by a variable $\lambda$, scaled by probability that initial advice was biased, given an adviser’s current and prior actions. We interpret this as self-image concerns within a dual-self model because our experimental setup tried to minimize social signaling concerns. However,

Figure A.1. Adviser-types and behavior in the $(\kappa, \lambda)$-plane

Note: a) $(\kappa, \lambda)$-values which imply persistent bias in BONUS for own choices (in rectangle) and b) for recommendations (in pentagon if own choices have diagnostic value, in triangle or pentagon above diagonal if not).
for the main effect we present and the above theoretical mechanism behind it, the source of image concerns is irrelevant – social image concerns can therefore have the same effect. The only crucial feature is that the party whose image of the adviser matters to him can observe the history of his actions.

It also deserves discussion that we model image concerns as being perceived as corrupted in past advice. Accordingly, such concerns do not play a role in the initial recommendation when there is no past behavior. There are two main reasons which entertain this assumption: The first is that in BONUS, the possibility to reveal oneself as an adviser who has given biased advice only occurs when the bonus is removed and one acts inconsistently afterwards. However, when advice was first given in R1, advisers in our experiment did not know that there would be a possibility to act inconsistently. Similarly, in many of the real-world settings we have in mind, advisers give biased advice until, through some unexpected exogenous events such as a scandal and resulting new regulations, this environment is reshaped by the removal of the conflict of interest. The second main reason in support of image concerns which are backward-looking are the results from ANTICIPATE: In this treatment, advisers were explicitly made aware of future situations which can raise image concerns. While we do observe that this intervention affect behavior, the way in which it does so is in line with the prediction which assumes that advisers do not factor in backward-looking image concerns when they form a plan ex-ante. However, the data is inconsistent with predictions assuming a forward-looking component for image concerns.
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 \((\kappa, \lambda)\). 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 \((\kappa, \lambda) \in \mathbb{R}_+^2 \), implies
\[
K'(x) = \int_0^\infty j(x, \lambda) d\lambda > 0.
\]
Repeating this for \( \Lambda \) proves part a). Part b) can be proven analogously, as for any \( x > 0 \)
\[\Lambda(y \mid x) = \Pr[\lambda \leq y \mid \kappa \leq x] = \left(\int_0^y \int_0^x j(\kappa, \lambda) d\kappa d\lambda\right) / \left(\int_0^\infty \int_0^x j(\kappa, \lambda) d\kappa d\lambda\right) \]
\[\Rightarrow \Lambda'(y \mid x) = \left(\int_0^y j(\kappa, y) d\kappa\right) / \left(\int_0^\infty j(\kappa, \lambda) d\kappa d\lambda\right) > 0.\]

For part c), rewrite the conditions \( \kappa \in [0, x] \) and \( \kappa/\lambda \leq z \) as \( \kappa \in \{0, \min\{x, z\}\} \). 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
\[
R(z \mid x, y) = \left(\int_y^\infty \int_0^{\min\{x, z\}} j(\kappa, \lambda) d\kappa d\lambda\right) / \left(\int_0^\infty \int_0^{x} j(\kappa, \lambda) d\kappa d\lambda\right)
\]
If \( x > z\lambda \), the partial derivative of the above w.r.t. \( z \) is then given by
\[
R'(z \mid x, y) = \left(\int_y^\infty \lambda \cdot j(z\lambda, \lambda) d\lambda\right) / \left(\int_y^\infty j(\kappa, \lambda) d\kappa d\lambda\right)
\]
and strictly positive by full support of \( j \). If \( x \leq z\lambda \), this derivative is zero so that \( R'(z \mid x, y) \geq 0 \) always holds. \( \Box \)

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 \( \tau_0 \) of type-1-advisers choose for themselves such that \( o \in \mathcal{B} \). Similarly, let \( \pi_o \) denote the share of type-3-advisers who choose for themselves \( o \in \mathcal{B} \). The following posteriors then emerge:

\[
\Pr[\theta = 3 \mid o \in \mathcal{B}, r_1 \in \mathcal{B}] = \frac{\pi_o \cdot \phi_3}{\tau_0 \cdot \phi_1 + \pi_o \cdot \phi_3}
\]
\[
\Pr[\theta = 3 \mid o \in \mathcal{N}, r_1 \in \mathcal{B}] = \frac{(1 - \pi_o) \cdot \phi_3}{(1 - \tau_o) \cdot \phi_1 + (1 - \pi_o) \cdot \phi_3}
\]

It is easily verified that the latter posterior is weakly larger than the former if and only if \( \tau_0 \geq \pi_o \). If this condition applies, then it holds for type-1-advisers (for whom \( o^* = o_B^\theta \)) 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 \( o = o^* \in B \). 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 \( o = o^* = r^* = r_1 \in B \). Type-3-advisers can thus pool with type-1s by choosing consistently from \( B \), i.e., \( o = r_1 \in B \), even though for them \( o^* \in \mathcal{N} \). They then choose their constrained optimum \( o^*_B \in B \). If they do not choose consistently they can choose their preferred option \( o^* \in \mathcal{N} \) but will then reveal themselves as corruptible, i.e., as type-3-advisers. Using (26) and the assumption that in case of indifference they choose \( o^* \), this means that type-3-advisers pool if the following holds:

\[
\begin{aligned}
& u(v(o^*)) - \lambda < u(v(o^*_B)) - \lambda \cdot \frac{\pi_o \phi_3}{\phi_1 + \pi_o \phi_3} \iff \lambda > \frac{(u(v(o^*)) - u(v(o^*_B))) \cdot \left( \frac{\phi_1 + \pi_o \phi_3}{\phi_1} \right)}{1 - \pi_o - \Lambda \left( (u(v(o^*)) - u(v(o^*_B))) \cdot \left( \frac{\phi_1 + \pi_o \phi_3}{\phi_1} \right) \right)};
\end{aligned}
\]

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

\[
1 - \pi_o = \Lambda \left( (u(v(o^*)) - u(v(o^*_B))) \cdot \left( \frac{\phi_1 + \pi_o \phi_3}{\phi_1} \right) \right).
\]

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 \( \pi_o \) 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 (26) is strictly larger than (27). Since for type-3-advisers \( o^* = o^*_B \) holds, they then choose their preferred choice as

\[
\begin{aligned}
& u(v(o^*)) - \lambda \cdot \Pr[\theta = 3 \mid o^*, r_1 \in B] > u(v(o^*_B)) - \lambda \cdot \Pr[\theta = 3 \mid o^*_B, r_1 \in B].
\end{aligned}
\]

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 \( r_1 \in B \), it holds that \( \Pr[\theta = 3 \mid r_1 \in \mathcal{N}, o, r_2] = 0 \) and type-2s therefore maximize \( U(o \mid r_1 \in \mathcal{N}, r^*) = -\kappa \cdot \mathbb{I}[r_2 \neq r^*] \) 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 \( 0 \) (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 \( r_2 \in C \). They then maximize \( U(r_2 \mid r_1, r_2, r^*) = -\kappa \cdot \mathbb{I}[r_2 \neq r^*] - \lambda \cdot \Pr[\theta = 3 \mid r_2, r_1, o] \) by recommending \( r_2 = r^* \in \mathcal{N} \). Therefore, they do not re-recommend their initial, biased recommendation \( r_1 \in B \).

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 \( o = r_1 \in B \). Accordingly, hitherto unrevealed type-3-advisers can continue to pool with type-1-advisers. Denote with \( \tau_{r_2} \) the share of type-1-advisers who recommend \( r_2 \in B \) and with \( \pi_{r_2} \) the share of type-3-advisers who recommend \( r_2 \in B \). This yields the following posteriors, conditional on not having previously revealed oneself (i.e., that \( o = r_1 \) holds):

\[
\Pr[\theta = 3 \mid r_2 \in B, r_1 \in B] = \frac{\pi_{r_2} \cdot \phi_3}{\tau_{r_2} \phi_1 + \pi_{r_2} \cdot \phi_3}.
\]

(28)
Posterior (29) is weakly larger than (28) if and only if $\tau_2 \geq \tau_1$. If this condition holds, the payoff for type-1-advisers with $r^* \in B$ from re-recommending $r^*$ in $R2$ is always strictly larger than from recommending $r_2' \in N$:

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

Thus, the only equilibrium with $\tau_2 \geq \tau_1$ obeys $\tau_2 = 1$. Hitherto unrevealed type-3-advisers who want to pool with type-1s have to choose analogously, i.e., $r_2 = r_1 \in B$, even though this is not their appropriate choice because for them, $r^* \in N$ holds. This allows them to not reveal themselves as corruptible so that their image costs $\lambda$ are discounted by $\Pr[v \mid \tau \geq 1 \in B, \phi = r_1 \in B]$. For this, they experience costs $\kappa$ of recommending something they do not consider appropriate. Plugging in the above posteriors with $\tau_2 = 1$ yields

$$-\lambda < \kappa < \lambda\cdot \frac{\phi_1}{\phi_1 + \tau_2 \cdot \pi_3^o} \Leftrightarrow \kappa < \lambda\cdot \frac{\phi_1}{\phi_1 + \tau_2 \cdot \pi_3^o}$$

as a condition for hitherto unrevealed type-3-advisers to continue pooling with type-1s. Note that Lemma 1b implies that for every $\kappa$ multiplied with some factor, there is a mass of advisers with sufficiently high $\lambda$, i.e., with $\lambda > \kappa(\phi_1 + \pi_2 \cdot \pi_3^o)/\phi_1$. Therefore, $\tau_2 = 0$ cannot be true. Also, the limit on $\kappa/\lambda$ 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 $\lambda > (u(v(o^*)) - u(v(o^o))) \cdot (\phi_1 + \pi_2 \cdot \pi_3^o)/\phi_1$). Thus, the share of type-3-advisers who continue to pool, denoted by $\tau_2$, is determined by the solution to

$$\pi_{\tau_2} = R \left( \frac{\phi_1}{\phi_1 + \pi_2 \cdot \pi_3^o} \mid u(b), (u(v(o^*)) - u(v(o^o))) \cdot (\phi_1 + \pi_2 \cdot \pi_3^o)/\phi_1 \right).$$

A solution $\pi_{\tau_2} = 0$ has been ruled out above. By Lemma 1c the above RHS is non-increasing in $\pi_{\tau_2}$ 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 $\pi_{\tau_2} \in (0,1]$. We exclude equilibria with $\tau_2 < \tau_1$ in a similar fashion as in the proof of Lemma 1. With $\tau_2 < \tau_1$, the posterior (28) is larger than (29). For hitherto unrevealed type-3-advisers with $r^* \in N$ it thus holds that

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

and they all recommend the choice $r^* \in N$ and thereby reveal themselves. This implies $\tau_2 = 0$ and thus contradicts an equilibrium with $\tau_2 \leq \tau_1$.

Recall from Lemma 1 and its proof that when what is considered appropriate advice is independent from own choices, $\phi$ 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 $\tau_2^o = 1$. The above reasoning can then be repeated with this parameter choice when the RHS in (30) is replaced by $R (\phi_1/(\phi_1 + \pi_2 \cdot \pi_3^o) \mid u(b), 0)$ 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 $\lambda$). The qualitative results, however, remain unchanged.

**Proof of Proposition 3**

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

**Lemma 2. Advisers with $r^* \in B$ prefer BBB over BBN (i.e., (2) over (3)) while there is a positive share of advisers with $r^* \in N$ who prefer BBB over BBN (i.e., (10) over (11)) and there is a positive share of such advisers who prefer BBN over BBB (i.e., (11) over (10)). Furthermore, $P_{BBN} = 1$.**
Proof. The relevant posteriors are given below, where \( \pi_o, \pi_{r_2}, \tau_o, \text{ and } \tau_{r_2} \) are defined as in the proof of Proposition 1 and 2:

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

Posterior (32) is weakly larger than (31) if and only if \( \tau_{r_2} \geq \pi_{r_2} \). If this condition holds, the payoff for advisers with \( r^* \in B \) from re-recommending \( r^* \) in R2 is always strictly larger than from recommending \( r_2 \in N \), i.e.,

\[-\lambda \cdot P_{BBB} > -\kappa - \lambda \cdot P_{BBN}.
\]

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

To exclude equilibria with \( \tau_{r_2} < \pi_{r_2} \), suppose to the contrary that this condition held. Then the posterior (32) is smaller than (31). Those advisers with \( r^* \in N \) who prefer BBN over BBB must then have

\[-\lambda - \lambda \cdot P_{BBB} < -\lambda \cdot P_{BBN}
\]

which means that they all recommend from \( N \) 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

\[P_{BBB} = \frac{\pi_{r_2} \cdot \pi_o \phi_3}{\tau_o \phi_1 + \pi_{r_2} \cdot \pi_o \phi_3} \quad \text{and} \quad P_{BBN} = 1.
\]

Therefore, advisers with \( r^* \in N \) choose BBB if and only if

\[-\kappa - \lambda \cdot P_{BBB} > -\lambda \cdot P_{BBN} \quad \Leftrightarrow \quad \kappa < \lambda \cdot \frac{\tau_o \phi_1}{\tau_o \phi_1 + \pi_{r_2} \cdot \pi_o \phi_3}.
\]

For advisers with \( r^* \in N \) who plan to recommend \( r_1 \in B \) and to choose \( o \in B \) it then follows from Lemma 1c that there is a positive mass of them who choose BBB and a positive mass who choose 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 \( r^* \in B \) prefer BBN over BNN (i.e. (19) over (21)), while there is a positive share of advisers with \( r^* \in N \) who prefer BBN over BNN (i.e., (12) over (11)) and a positive share of such advisers who prefer BNN over BBN (i.e., (11) over (12)). Furthermore, \( P_{BNN} = P_{BN} = 1 \).

**Proof.** The relevant posteriors are given by

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

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 \( r^* \in B \) prefer BBN over BNN since

\[u(o^*) - \lambda \cdot P_{BB} - \lambda \cdot P_{BBN} > u(o_B^*) - \lambda \cdot P_{BN} - \lambda \cdot P_{BNN}.
\]
The above follows from the fact that $P_{BBN} = P_{BNN} = 1$. To see this, consider these two posteriors:

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

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

It can be easily verified that $P_{BBN} \leq P_{BNN}$ if $\tau_o \geq \pi_o$. From Lemma 2, we get that $P_{BNN} = 1$ which then implies $P_{BBN} = 1$. As (34) holds, when then get $\tau_o = 1 \geq \pi_o$ for all equilibria with $\tau_o \geq \pi_o$. This also means $P_{BN} = 1$.

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

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

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

$$u(o^*_o) - \lambda \cdot P_{BB} > u(o^*) - \lambda \iff \lambda > \frac{u(o^*) - u(o^*_o)}{1 - P_{BB}}.
$$

(35)

Given that $u(o^*) > u(o^*_o)$ and $P_{BB} < 1$, we know that the RHS of the above inequality is strictly positive. According to full support assumption on the distribution of $\kappa$ and $\lambda$, the probability that $\lambda$ 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 $BNN$ over $BBN$ is strictly positive and less than one, and that the mass of advisers with $r^* \in \mathcal{N}$ who prefer $BBN$ over $BNN$ 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 plan never features $o \in \mathcal{N}$:

**Lemma 4.** In every equilibrium, advisers with $r^* \in \mathcal{B}$ prefer $BBB$ over $BNN$ and $BNN$ (i.e. (3) over (4) and (5)). Furthermore, $P_{BNB} = 1$.

*Proof.* Using $P_{BBN} = P_{BNN} = 1$, one gets from (4) that advisers with $r^* \in \mathcal{B}$ who choose $BNN$ have a payoff of $u(b) + u(o^*_o) - \lambda - \kappa - \lambda$. This is lower than (2), the payoff from choosing $BBB$ which is given by $u(b) + u(o^*) - \lambda \cdot P_{BB} - \lambda \cdot P_{BBB}$. In a similar manner, (5), the payoff from choosing $BNB$, becomes $u(b) + u(o^*_o) - \lambda - \lambda$ which is also strictly lower than (2). 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:

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

□

**Lemma 5.** In every equilibrium, advisers with $r^* \in \mathcal{N}$ prefer $BNN$ over $BNB$ (i.e., (12) over (13)).

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

$$-\lambda \cdot P_{BBN} \geq -\kappa - \lambda \cdot P_{BBB}$$

which is always true given that $P_{BBN} = P_{BNN} = 1$.

□

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

18
**Lemma 6.** In every equilibrium, a strictly positive mass of advisers with \( r^* \in \mathcal{N} \) prefer \( \text{BBB} \) over \( \text{BNN} \), whereas a strictly positive mass of these advisers prefer \( \text{BBN} \) over \( \text{BBB} \).

**Proof.** Advisers with \( r^* \in \mathcal{N} \) prefer \( \text{BBN} \) over \( \text{BBB} \) if and only if (12) is no less than (10), i.e.,

\[
u(o^*) - 2\lambda \geq \nu(o_B) - \lambda \cdot P_{BB} - \kappa - \lambda \cdot P_{BBB}
\]

\[
\Leftrightarrow \quad \kappa \geq \lambda \cdot [1 - P_{BB} + 1 - P_{BBB}] - [\nu(o^*) - \nu(o_B)].
\]

(36)

For \( \lambda \) 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 \((\kappa, \lambda)\) 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 \( r^* \in B \) who plan to choose \( r_1 \in \mathcal{N} \) prefer \( \text{NNB} \) over \( \text{NNN} \), \( \text{BNN} \), and \( \text{NNN} \) (i.e., (9) over (7), (8), and (6)). Advisers with \( r^* \in \mathcal{N} \) who plan to choose \( r_1 \in \mathcal{N} \) prefer \( \text{NNN} \) over \( \text{NNB} \), \( \text{BNN} \), and \( \text{NNN} \) (i.e., (14) over (15), (16), and (17)).

**Proof.** Lemma 7 follows from two comparisons: First, comparing (9) to (6), (7), and (8), respectively and second, comparing (14) to (15), (16), and (17), respectively.

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

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

holds. Therefore, the percentage of advisers with \( r^* \in B \) who choose \( r_1 \in B \) is given by

\[
\tilde{r}_{r_1}^* = \int_0^\infty \int_0^{\max(0, \lambda \cdot (P_{BB} + P_{BBB}) - \nu(b))} j(\kappa, \lambda) \, d\lambda \, d\kappa = \int_0^\infty \int_0^\infty j(\kappa, \lambda) \, d\lambda \, d\kappa = 1.
\]

The above proves the first part of the proposition. Advisers with \( r^* \in \mathcal{N} \), on the other hand, may prefer \( \text{BBB}, \text{BBN}, \) or \( \text{BBN} \) if they plan to recommend \( r_1 \in B \). In case they plan to recommend \( r_1 \in \mathcal{N} \), the previous results show that they only do so in the sequence \( \mathcal{N} \). For advisers with \( r^* \in \mathcal{N} \), the share of them who recommend \( r_1 \in B \) is therefore determined by comparing each of the plans \( \text{BBB}, \text{BBN}, \) and \( \text{BBN} \), to \( \text{NNN} \) if such a plan is most preferred among the plans which feature \( r_1 \in B \). For this, it is convenient to denote with \( \Pi_1, \Pi_2, \) and \( \Pi_3 \) the three partitions which divide the mass of advisers with \( r^* \in \mathcal{N} \) and who recommend \( r_1 \in B \) in equilibrium:

\( \Pi_1 \): Advisers who prefer \( \text{BBB} \) over the rest of plans which feature \( r_1 \in B \) and over \( \mathcal{N} \).

\( \Pi_2 \): Advisers who prefer \( \text{BBN} \) over the rest of plans which feature \( r_1 \in B \) and over \( \mathcal{N} \).

\( \Pi_3 \): Advisers who prefer \( \text{BBN} \) over the rest of plans which feature \( r_1 \in B \) and over \( \mathcal{N} \).

For advisers in \( \Pi_1 \), (33) and (35) must be satisfied for \( \text{BBB} \) being preferred over the rest of plans which feature \( r_1 \in B \). In addition, the following condition must hold to ensure that \( \text{BBB} \) is preferred over \( \mathcal{N} \), i.e., that (10) is larger than (14):

\[
\kappa < \frac{1}{2} [\nu(b) - \nu(o^*) - \nu(o_B)] - \lambda \cdot (P_{BB} + P_{BBB}).
\]

(37)

In consequence, \( \Pi_1 \) can also be defined via the restriction put on the \((\kappa, \lambda)\)-values of the advisers:

\[
\Pi_1 \equiv \{(\kappa, \lambda) \text{ s.t. conditions (33), (35), and (37) are satisfied}\}
\]
Similarly, for advisers in \( \Pi_2 \), (35) and the opposite of (33) must be satisfied to ensure that \( \text{BBN} \) is preferred over the rest of plans which feature \( r_1 \in B \). In addition, the following condition must hold to ensure that \( \text{BBN} \) is preferred over \( \mathcal{NNN} \), i.e., that (11) is larger than (14):

\[
\kappa < u(b) - [u(o^*) - u(o_{rb}^*)] - \lambda \cdot (P_{BB} + 1). \tag{38}
\]

This then allows to (re-)define \( \Pi_2 \) as follows:

\[
\Pi_2 \equiv \{ (\kappa, \lambda) \text{ s.t. condition (35), condition (38), and the opposite of condition (33) are satisfied} \}
\]

For advisers in \( \Pi_3 \), (36) and the opposite of (33) must be satisfied to ensure that \( \text{BNN} \) is preferred over the rest of plans which feature \( r_1 \in B \). In addition, the following condition must hold to ensure that \( \text{BNN} \) is preferred over \( \mathcal{NNN} \), i.e., that (12) is larger than (14):

\[
\kappa < u(b) - 2 \cdot \lambda. \tag{39}
\]

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

\[
\Pi_3 \equiv \{ (\kappa, \lambda) \text{ s.t. condition (36), condition (39), and the opposite of condition (35) are satisfied} \}
\]

For ease of exposition, denote by \( k_{33}(\lambda) \) the RHS of (33), \( k_{36}(\lambda) \) the RHS of (36), \( k_{37}(\lambda) \) the RHS of (37), \( k_{38}(\lambda) \) the RHS of (38), \( k_{39}(\lambda) \) the RHS of (39), and \( l_{35} \) the RHS of (35). We can then compute the share of advisers within these three partitions as the share of the total population of advisers as follows:

\[
\tilde{\pi}^{*}_{r_1} = \sum_{t=1}^{3} \int_{\Pi_t} j(\kappa, \lambda) d\kappa d\lambda
\]

\[
= \int_{l_{35}}^{\infty} \int_0^{\min\{k_{33}(\lambda), k_{37}(\lambda)\}} j(\kappa, \lambda) d\kappa d\lambda + \int_{l_{35}}^{\infty} \int_{k_{33}(\lambda)}^{k_{39}(\lambda)} j(\kappa, \lambda) d\kappa d\lambda + \int_{l_{35}}^{\max\{k_{33}(\lambda), k_{36}(\lambda)\}} j(\kappa, \lambda) d\kappa d\lambda
\]

\[
< \int_{l_{35}}^{\infty} \int_0^{u(b)} j(\kappa, \lambda) d\kappa d\lambda + \int_{l_{35}}^{\max\{k_{33}(\lambda), k_{36}(\lambda)\}} \int_0^{u(b)} j(\kappa, \lambda) d\kappa d\lambda
\]

\[
= \int_{l_{35}}^{\infty} \int_0^{u(b)} j(\kappa, \lambda) d\kappa d\lambda = K(u(b)).
\]

The inequalities are because the conditions (37), (38), and (39) restrict \( \kappa \) – and therefore, the respective upper limits on it in the above integrals – to be strictly less than \( u(b) \). Also, \( l_{35} > 0, k_{33}(\lambda) \geq 0 \) holds for any \( \lambda \geq 0 \). Hence, the mass of advisers with \( r^* \in \mathcal{N} \) who recommend \( r_1 \in B \) is lower than \( K(u(b)) \) which proves the second part of the proposition. \( \square \)

**Proof of Proposition 4**

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 (18) through (25).

**Lemma 8.** If advisers do not anticipate image costs, those with \( r^* \in B \) and \( o^* \in B \) prefer \( \text{BBB} \) among all possible plans of actions.
Lemma 9. If advisers do not anticipate image costs, all advisers with $r^* \in B$ and $o^* \in B$ choose either $BNN$ or $NNN$. The share of these advisers who prefer the former over the latter is given by $K(u(b))$.

Proposition 4 then follows immediately from the above lemmas.
Figure C.1. Full distributions of advisers' actions (rows) over treatments (columns)

Note: Bars depict standard errors.
<table>
<thead>
<tr>
<th></th>
<th>NO BONUS</th>
<th>BONUS</th>
<th>ANTICIPATE</th>
<th>OVERALL</th>
<th>KW/χ²-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>mean</td>
<td>s.d.</td>
<td>p-value</td>
</tr>
<tr>
<td>Age</td>
<td>24.82</td>
<td>8.002</td>
<td>23.208</td>
<td>5.411</td>
<td>0.110</td>
</tr>
<tr>
<td>Male</td>
<td>0.451</td>
<td>0.503</td>
<td>0.354</td>
<td>0.483</td>
<td>0.536</td>
</tr>
<tr>
<td>Region of origin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.049</td>
</tr>
<tr>
<td>Europe/N. America/Australia/NZ</td>
<td>0.353</td>
<td>0.483</td>
<td>0.333</td>
<td>0.476</td>
<td>0.376</td>
</tr>
<tr>
<td>Asia</td>
<td>0.608</td>
<td>0.493</td>
<td>0.646</td>
<td>0.483</td>
<td>0.557</td>
</tr>
<tr>
<td>Other</td>
<td>0.039</td>
<td>0.196</td>
<td>0.021</td>
<td>0.144</td>
<td>0.067</td>
</tr>
<tr>
<td>Degree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.460</td>
</tr>
<tr>
<td>Bachelor</td>
<td>0.608</td>
<td>0.493</td>
<td>0.500</td>
<td>0.505</td>
<td>0.544</td>
</tr>
<tr>
<td>Master</td>
<td>0.353</td>
<td>0.483</td>
<td>0.479</td>
<td>0.505</td>
<td>0.430</td>
</tr>
<tr>
<td>PhD</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Other postgraduate</td>
<td>0.000</td>
<td>0.000</td>
<td>0.021</td>
<td>0.144</td>
<td>0.007</td>
</tr>
<tr>
<td>None</td>
<td>0.039</td>
<td>0.196</td>
<td>0.000</td>
<td>0.000</td>
<td>0.020</td>
</tr>
<tr>
<td>Study subject</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.294</td>
</tr>
<tr>
<td>Economics/Business/Finance</td>
<td>0.216</td>
<td>0.415</td>
<td>0.375</td>
<td>0.489</td>
<td>0.315</td>
</tr>
<tr>
<td>Other social sciences</td>
<td>0.353</td>
<td>0.483</td>
<td>0.229</td>
<td>0.425</td>
<td>0.295</td>
</tr>
<tr>
<td>Psychology</td>
<td>0.059</td>
<td>0.238</td>
<td>0.021</td>
<td>0.144</td>
<td>0.054</td>
</tr>
<tr>
<td>Public administration</td>
<td>0.039</td>
<td>0.196</td>
<td>0.063</td>
<td>0.245</td>
<td>0.040</td>
</tr>
<tr>
<td>Math/Sciences/Engineering</td>
<td>0.157</td>
<td>0.367</td>
<td>0.083</td>
<td>0.279</td>
<td>0.087</td>
</tr>
<tr>
<td>Arts or Humanities</td>
<td>0.157</td>
<td>0.367</td>
<td>0.146</td>
<td>0.357</td>
<td>0.148</td>
</tr>
<tr>
<td>Other</td>
<td>0.020</td>
<td>0.140</td>
<td>0.083</td>
<td>0.279</td>
<td>0.060</td>
</tr>
<tr>
<td>Monthly budget (in GBP)</td>
<td>606.3</td>
<td>450.7</td>
<td>640.0</td>
<td>563.8</td>
<td>0.088</td>
</tr>
</tbody>
</table>

Number of observations | 51 | 48 | 50 | 149

Note: The rightmost column provides p-values for the null of equality between the three treatments (Kruskal-Wallis tests for the variables age and budget; χ²-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
Screen 2a: The clients’ choice situation and information about the bonus.

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 later 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 £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, 2, 3: client earns an amount of £12.00.
  - Die shows 4, 5, or 6: client earns £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 £12.00.
  - Die shows 3, 4, or 5: client earns an amount of £20.00 when the coin shows “Heads’, or nothing when the coin shows “Tails”.

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

If you choose a different option for yourself, and then another client, both the client and you will follow the same guidelines for your respective choices.
Please look now at your 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.
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.

<table>
<thead>
<tr>
<th>Die equal to…</th>
<th>Option A is chosen</th>
<th>Option B is chosen</th>
<th>Option C is chosen</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 or 2</td>
<td>lottery: GBP 20 or 0</td>
<td>GBP 12</td>
<td>GBP 12</td>
</tr>
<tr>
<td>3 or 4</td>
<td>lottery: GBP 20 or 0</td>
<td>lottery: GBP 20 or 0</td>
<td>GBP 8</td>
</tr>
<tr>
<td>5 or 6</td>
<td>lottery: GBP 20 or 0</td>
<td>lottery: GBP 20 or 0</td>
<td>lottery: GBP 20 or 0</td>
</tr>
</tbody>
</table>

**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 is 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 sentence.
- 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.

---

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 envelopes that contain these choices.

The following happens if your envelope is randomly chosen:

- We will read your numerical 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.

---

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 be asked to write down your recommendation and put it in an envelope as with your previous recommendation and your own choices. All the envelopes will be randomly chosen from those that contain these choices and one will be selected to go to the 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, address 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.

Screen 5: Instructions for giving the second recommendation R2

Some last questions
Before finishing the experiment, we would like to ask some additional questions about you. All answers will be recorded 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 willing  very unwilling

Please choose your gender:
• male
• female

What is your age (in years)?

Which of the following best describes the region you are from?

Which of the following describes your most recent field of study best?

Which of the following describes your highest degree you are holding or pursuing?

What is the monthly budget (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.

Screen 6: Exit questionnaire