Center for Nonlinear Dynamics in Economics and Finance

CeNDEF Working Paper 26–07

Valeria Burdea and Anita Kopányi–Peuker

Disclosure of Conflicts of Interest in Long-Run Relationships

November 2026

Abstract

In many professional contexts, experts advise less-informed decision makers while holding potential conflicts of interest (COI). To mitigate the risks of biased advice, institutions increasingly mandate COI disclosure, yet prior theoretical and experimental work suggests that such mandatory disclosure can backfire in one-shot interactions. Why, then, is disclosure so prevalent in practice? We study two often-overlooked features of real-world advisory relationships—repeated interaction and voluntary disclosure—through a laboratory experiment. We use a setting where an advisor, whose preferences are either aligned or misaligned with the client's, privately observes a state and sends a cheap-talk message about it before the client chooses an action. We vary whether the advisor's preferences are hidden from the client (no disclosure), disclosed by default (mandatory), or disclosed at the advisor's discretion (voluntary), and whether pairs are randomly rematched or interact as fixed partners (short-term vs long-term partner matching). We find that mandatory disclosure benefits clients in short-term interactions, without hurting advisors on average. Under long-term matching, only voluntary disclosure improves clients' payoffs relative to no disclosure. Misaligned advisors who voluntarily disclose their COI exaggerate less, and their advice is followed more often, particularly in long-term relationships. Clients, in turn, treat non-disclosure as a bad signal. Our results suggest that disclosure, and voluntary disclosure in particular, can foster trust and improve decision accuracy.

University of Amsterdam
Amsterdam School of Economics

Disclosure of Conflicts of Interest in Long-Run Relationships*

Valeria Burdea† and Anita Kopányi-Peuker‡

November 2, 2026

Abstract

In many professional contexts, experts advise less-informed decision makers while holding potential conflicts of interest (COI). To mitigate the risks of biased advice, institutions increasingly mandate COI disclosure, yet prior theoretical and experimental work suggests that such mandatory disclosure can backfire in one-shot interactions. Why, then, is disclosure so prevalent in practice? We study two often-overlooked features of real-world advisory relationships—repeated interaction and voluntary disclosure—through a laboratory experiment. We use a setting where an advisor, whose preferences are either aligned or misaligned with the client's, privately observes a state and sends a cheap-talk message about it before the client chooses an action. We vary whether the advisor's preferences are hidden from the client (no disclosure), disclosed by default (mandatory), or disclosed at the advisor's discretion (voluntary), and whether pairs are randomly rematched or interact as fixed partners (short-term vs. long-term partner matching). We find that mandatory disclosure benefits clients in short-term interactions, without hurting advisors on average. Under long-term matching, only voluntary disclosure improves clients' payoffs relative to no disclosure. Misaligned advisors who voluntarily disclose their COI exaggerate less, and their advice is followed more often, particularly in long-term relationships. Clients, in turn, treat non-disclosure as a bad signal. Our results suggest that disclosure, and voluntary disclosure in particular, can foster trust and improve decision accuracy.

Keywords: Conflict of interests, Mandatory disclosure, Voluntary disclosure, Laboratory experiment
JEL codes: C72, C92, D82, D83

*We are grateful for comments from seminar audiences at LMU Munich, ISER at the University of Osaka, the University of Zurich, the University of Groningen, the University of Stavanger, the Corvinus University of Budapest and the University of Amsterdam, and from conference participants at the Experimental Finance conference in Maastricht, the RBFC conference in Amsterdam, IMEBESS in Lisbon, TIBER in Tilburg. IRB approval is received from Radboud University (approval number: EACLM-LT-048) with which Kopányi-Peuker was affiliated when running the experiment. Financial support from the Deutsche Forschungsgemeinschaft through CRC TRR 190 (project number 280092119) and Radboud University is gratefully acknowledged.

†Ludwig Maximilian University of Munich
‡University of Amsterdam, Tinbergen Institute

1 Introduction

People often consult professionals about important decisions on which they lack expertise, for example about their finances, their health, or the sale of their house. In these situations, clients typically assume that their advisors care about them and recommend the option that is best for them. However, this is not always the case: the advisor may face a potential conflict of interest (henceforth COI). Think of a financial advisor who receives a commission if the client chooses a certain investment, or a doctor who receives financial support from a pharmaceutical company. A common policy response is to require advisors to disclose their potential conflicts of interest,[1] even though theoretical (Li and Madarász, 2008) and experimental (Cain et al., 2005) research shows that such policies can backfire on the very people they are meant to protect: the clients. Mandatory disclosure can lead advisors to exaggerate their advice, anticipating that clients will discount it. Moreover, research shows that clients can be either overly skeptical (Li and Madarász, 2008) or not skeptical enough (Cain et al., 2005) when responding to the disclosure of a COI. Why, then, is mandatory disclosure so common in practice?

In this paper, we address this question by examining two often-overlooked features of advisory relationships. First, we study the role of repeated interaction. Real-world advisor–client relationships are often long-lasting, whereas most research focuses on one-shot interactions. People do not change their doctor after each visit; they often prefer to stay with the same physician if possible. Long-lasting relationships allow both parties to learn about each other's behavior and to build trust. Furthermore, reputation and the threat of punishment can become credible forces in such relationships.

Second, we ask whether another form of disclosure, voluntary disclosure, can also protect clients and enhance welfare. Voluntary disclosure cannot be mandated, but advisors can choose whether to share such sensitive information with their clients. In our setting, voluntary disclosure should not add value beyond mandatory disclosure if all agents are rational and maximize their own payoffs. However, behavioral insights suggest that by voluntarily sharing information, advisors may build trust with their clients more easily. This, in turn, can improve advice-giving and advice-taking, potentially resulting in higher welfare.[2] Our study therefore addresses the following research questions: *Does mandatory disclosure help clients in long-term rather than short-term interactions? Can voluntary disclosure perform even better?*

To answer these questions, we design a between-subjects laboratory experiment based on a cheap-talk game (Crawford and Sobel, 1982). Participants are paired and assigned the role of either Advisor or Client. The Advisor observes the state of the world, which determines the Client's payoff, and can send a message about it. The Client observes only the message and then chooses an action. The closer the Client's action is to the realized state, the higher the Client's payoff. The Advisor is one of two types: her preferences are either perfectly aligned with the Client's or misaligned. Participants play this cheap-talk game repeatedly with fixed roles, while the

[1] Examples of such policies include the Physician Payments Sunshine Act (2010), which requires disclosure of financial transfers between manufacturers and health care providers; the amendments to Form ADV by the U.S. Securities and Exchange Commission (2010), which require advisors to disclose their potential COI; and the Federal Trade Commission's *.com Disclosures* guidance (2013), which aims to improve disclosure in online marketing, for example by bloggers.

[2] Kopányi-Peuker et al. (2017) show that cooperation in a prisoner's dilemma increases when players can choose their own potential punishment levels, but not when the same punishment levels are exogenously imposed on them.

Advisor's preferences and the state of the world are redrawn in every period. Our six treatments vary along two dimensions. First, to test the effect of the length of interaction, we implement either a *short-term* or a *partner* matching. Second, to investigate the effect of different disclosure regimes, we implement three disclosure protocols: *no disclosure*, *mandatory disclosure*, and *voluntary disclosure*.

Our experiment shows that both Advisors and Clients respond to the treatment variations. Advisors with misaligned preferences often exaggerate in their messages by over-reporting the state. Partner matching reduces the frequency of exaggeration. Turning to the disclosure protocols, we find no evidence that mandatory disclosure licenses Advisors to lie compared to no disclosure, in either matching setting. However, voluntarily disclosing a COI reduces exaggeration and increases truth-telling, and this effect is especially pronounced under partner matching. Advisors whose preferences are aligned with the Client's almost always tell the truth.

For Clients, we distinguish two cases. When the Advisor's preferences are not disclosed, Clients interpret voluntary non-disclosure as a bad sign and discount the message more often than in the no-disclosure treatment. Partner interaction reduces this skepticism. When the Advisor's preferences are disclosed and misaligned, partner interaction increases trust in the advice under voluntary disclosure: Clients discount the message less often than under short-term matching. When preferences are known to be aligned, Clients almost always follow the message. These behavioral differences translate into differences in decision accuracy. Clients' actions are more often perfectly accurate (i.e., they match the realized state) when preferences are disclosed than when they are not. Partner interaction slightly increases accuracy, especially when misaligned preferences are voluntarily disclosed. In terms of welfare, Advisors are, on average, neither helped nor harmed by any of our treatment variations. Clients benefit from partner matching when disclosure is not possible, and they benefit from voluntary disclosure (compared to no disclosure) under both matching protocols. Mandatory disclosure helps Clients only under short-term matching.

These results show that, in our experiment, Advisors do not feel licensed to lie when their COI is known. As a result, mandatory disclosure has no harmful effects under short-term matching; in fact, Clients benefit from both types of disclosure in short-term interactions. Disclosure plays an important role in identifying Advisors without a COI, which improves the quality of advice and increases efficiency. Furthermore, repeated interaction, especially when combined with voluntary disclosure, helps build more trusting relationships in which Advisors transmit more information and Clients can use this information to increase their welfare without making Advisors worse off.

Our paper contributes to the literature on disclosure protocols for COI in several ways. First, we are the first to study this problem in a setting in which we control Clients' beliefs about Advisors' preferences in all treatments, by explicitly informing them about the possible preferences and the probability of each type. Previous studies have mainly focused on a single type of advisor, varying the disclosure protocol when, for example, only biased advisors are present. Second, we investigate the effect of long-term relationships between advisors and clients under different disclosure protocols, whereas previous research has mainly focused on one-shot interactions.[3] Finally, we introduce voluntary disclosure of a COI, in which Advisors decide whether to reveal their type to their Client. To the best of our knowledge, voluntary disclosure has not yet been studied in isolation.[4]

[3] For an exception, see Koch and Schmidt (2010).

[4] Sah and Loewenstein (2014) couple the voluntary disclosure decision with the decision whether to accept a COI in the first place.

The paper is structured as follows. Section 2 describes related literature. Section 3 discusses the experimental game together with predictions and practicalities. Section 4 presents the results. Section 5 concludes. We relegate additional material (like experimental instructions and further results) to the Appendix.

2 Related literature

Our paper contributes to several strands of literature which we discuss in this section.

Disclosure of conflicts of interest. There are several studies investigating the effect of mandatory disclosure of conflict of interests. In a series of laboratory experiments Cain et al. demonstrated that mandatory disclosure can backfire (2005, 2011). In their main experiments advisors had more information about the object of the estimation (amount of money in a jar or the value of houses for sale) than their clients. Based on their personal estimates, advisors sent an advise to the clients who then formulated their own estimates which formed the base of the payment of both parties. These papers show that disclosing the presence of COI hurts clients, as clients do not discount the advice sufficiently. The outcome for the advisor was the same under both disclosure protocols. Later, Sah (2019) replicated the results of Cain et al. (2011) on the sale price of houses. Sah and Loewenstein (2014) also show that mandatory disclosure can hurt clients, but only when the COI is exogenously given for advisors (their Experiment 1). Their main question is the effect of the possibility of avoiding conflict. They find that if advisors can avoid a COI, and reject it, disclosure of preference alignment or COI helps clients (their Experiments 2 and 3). An important difference to our paper is that we investigate the effect of the disclosure protocols, and therefore we exogenously fix the advisors' preferences. Hence, none of the differences across our treatments can be attributed to fewer advisors taking up the conflict in the first place. In different settings Church and Kuang (2009), Ismayilov and Potters (2013), Chen and Richardson (2018) found that mandatory disclosure neither hurts nor helps. Closest to our interest in long-term relationships, Koch and Schmidt (2010) study disclosure when advisors can build a reputation, and find no harmful effect of mandatory disclosure in that case. In their design, however, an advisor is either always biased or always unbiased, whereas in our design the Advisor's preferences are redrawn every period and both types are present in all treatments.

Previous studies not only investigated the direct effect of mandatory disclosure on outcomes, but also researched the potential psychological forces behind it. On the advisor side, disclosure can give advisors a sense of *moral licensing*: having warned the client, advisors feel less responsible for the quality of their advice and exaggerate more (Cain et al., 2005, 2011; Loewenstein et al., 2011). This mechanism is related to the broader psychological literature on moral credentials, which shows that establishing one's moral standing can license subsequent behaviour that would otherwise appear questionable (Monin and Miller, 2001). Li and Madarász (2008) provide a strategic counterpart: advisors exaggerate more under disclosure because they anticipate that clients will discount their advice. On the client side, disclosure may decrease trust in the advice but at the same time increase the pressure to follow it. Clients may feel a "burden of disclosure": following the advice of an advisor who disclosed her conflict of interests is perceived as a socially desirable way of helping her (the *panhandler effect*; Sah et al., 2013). Moreover,

clients may fear that rejecting the advice signals that they suspect the advisor of being biased (*insinuation anxiety*; Sah et al., 2019). Notably, Sah et al. (2019) find that insinuation anxiety arises both when disclosure is voluntary and when it is required by law, but is reduced when the conflict is disclosed by an external source rather than by the advisor. Finally, disclosure can affect how clients perceive the advisor's character: Sah and Feiler (2020) show that disclosure leads to a "disclosure penalty" in trust even when advice is of high quality, but that disclosure can also act as an "altruistic signal" that increases trust when the advice runs counter to the advisor's self-interest (see also Loewenstein et al., 2012). Our design lets us assess whether such forces survive in a stylised strategic environment in which clients know the distribution of advisor types. In contrast to the moral-licensing prediction, we find that mandatory disclosure does not increase exaggeration. Consistent with a signalling role of transparency, voluntarily disclosed misaligned Advisors in partner relationships communicate more truthfully and receive more trust. This is in line with Kopányi-Peuker et al. (2017), who show that voluntarily making oneself vulnerable fosters cooperation, while the same vulnerability imposed exogenously does not.

Our Voluntary treatments relate to the literature on the voluntary disclosure of verifiable information. The unravelling argument of Grossman (1981) and Milgrom (1981) predicts full disclosure when information is verifiable and costless to disclose, because receivers interpret silence as bad news (see Dranove and Jin, 2010, for a review). Laboratory tests of this prediction find that senders with favourable information disclose, but disclosure is less than complete, and receivers are often insufficiently skeptical about non-disclosure (Jin et al., 2021); see also Hagenbach and Perez-Richet (2018) for experimental evidence on communication with evidence. In our setting, disclosure concerns the Advisor's preferences rather than the state, and it is combined with a cheap-talk message about the state. To the best of our knowledge, the only other study of voluntary disclosure of a conflict of interests is Sah and Loewenstein (2014), who couple the disclosure decision with the decision whether to accept the conflict in the first place.

Cheap talk. Our experimental setting builds on the canonical sender-receiver cheap-talk framework of Crawford and Sobel (1982), which has inspired a large theoretical (e.g., Sobel, 1985; Morgan and Stocken, 2003; Li and Madarász, 2008) and experimental literature (see Blume et al., 2020, for a survey). A robust finding across this literature is *overcommunication*: when there is a conflict of interests, senders reveal more information than predicted by the most informative equilibrium, and receivers rely on messages more than they should (Cai and Wang, 2006; Sánchez-Pagés and Vorsatz, 2007; Wang et al., 2010). Explanations for overcommunication include level-k reasoning (Cai and Wang, 2006; Wang et al., 2010; Kawagoe and Takizawa, 2009) and a preference for truth-telling. The latter is documented extensively in the literature on lying, which shows that many people are willing to forgo monetary gains rather than lie, and that lying aversion depends on the consequences of the lie for others and on how the lie is perceived (Gneezy, 2005; Fischbacher and Föllmi-Heusi, 2013; Gneezy et al., 2018; Abeler et al., 2019). We take this into account in our experiment and measure participants' lying preferences with a task based on Fischbacher and Föllmi-Heusi (2013), to better understand Advisors' disclosure and messaging decisions.

Two features distinguish our design from most cheap-talk experiments. First, the Client does not know whether the Advisor's interests are aligned with his own. This uncertainty about the sender's bias is central to theoretical

work on expert advice (Bénabou and Laroque, 1992; Morgan and Stocken, 2003; Li and Madarász, 2008). In these models, as in our Lemma 1, the credibility of a message depends on the receiver's belief about who sent it, and favourable messages (in our case, high messages) are pooled across sender types and are therefore discounted. Second, we study repeated interaction between the same Advisor and Client. Theory suggests that reputational concerns can sustain more informative communication when the sender's type is uncertain (Sobel, 1985; Bénabou and Laroque, 1992; Stocken, 2000), although reputation can also distort advice when an honest advisor wants to avoid appearing biased (Morris, 2001; Ely and Välimäki, 2003). Experimental evidence on repeated cheap talk is scarce. Wilson and Vespa (2020) find that in an indefinitely repeated cheap-talk game, outcomes are closer to babbling than in comparable one-shot settings, even though full revelation can be sustained in equilibrium; honest revelation emerges only when players can coordinate on an information rent for the sender. Our Partner treatments complement this evidence by showing how disclosure institutions interact with the shadow of the future.

Creedence goods. The advisor–client relationship we study shares important features with markets for credence goods, in which an expert knows better than the customer which good or service the customer needs, and the customer may never learn the quality of what was provided (Darby and Karni, 1973; Dulleck and Kerschbamer, 2006). Experimental work in this area studies how liability, verifiability, reputation and competition affect fraud and efficiency (Dulleck et al., 2011; Mimra et al., 2016); see Balafoutas and Kerschbamer (2020) for a review. Field evidence on financial advice shows that commission-based incentives can lead advisors to recommend products that are not in their clients' best interests (Mullainathan et al., 2012; Anagol et al., 2017), and theoretical work studies how commissions and their disclosure shape such advice (Inderst and Ottaviani, 2012). Similar to credence-goods markets, our Clients cannot verify the state after the interaction, so Advisors cannot be held accountable for individual messages, and repeated interaction is one of the few disciplining forces available. Unlike most credence-goods experiments, however, our Advisors do not choose prices or treatment plans: they influence outcomes only through costless messages, and their incentives are given by an exogenous conflict of interests rather than by a price structure. This allows us to isolate the role of communication and of information about the conflict itself.

Persuasion by conflicted advisors. Finally, our paper relates to experimental work on persuasion by conflicted advisors. In a high-stakes experiment on face-to-face insurance sales, de Meza et al. (2010) find that requiring sellers to disclose their commission and the policy's value for money has almost no effect on buyers' decisions; instead, buyers' general trust drives purchases, as trusting buyers are more likely to follow advice. Barron and Fries (2024) show that financial advisors with misaligned incentives construct self-interested narratives that shift investors' beliefs, and that this form of persuasion is difficult to protect against with standard policy interventions. In contrast to these richer settings, our Advisors can communicate only through a coarse cheap-talk message and Clients know the distribution of Advisor types, which lets us isolate how information about the conflict itself affects communication and trust. Credit rating agencies are a prominent real-world example of informed intermediaries with a conflict of interests, as they are typically paid by the issuers whose securities they rate. Theory shows that this can lead to rating inflation, and that reputational concerns may be insufficient to discipline rating agencies

(Mathis et al., 2009; Skreta and Veldkamp, 2009; Bolton et al., 2012). Rabanal and Rud (2018) study experimentally how competition between rating agencies affects their incentives to report truthfully, and Bayona et al. (2026) show in a laboratory experiment that potentially inflated ratings can still improve market outcomes, mainly because ratings help investors coordinate. Our setting abstracts from competition and market interaction, but similarly asks how transparency about the informed party's incentives and the possibility of reputation building affect the informativeness of advice.

3 Research design

This section presents the experimental design. In Section 3.1 we describe the setup of the game. In Section 3.2 we introduce our treatments, and in Section 3.3 we present the hypotheses. Section 3.4 describes the second part of the experiment, and Section 3.5 concludes with practical details about the experiment.

3.1 Experimental game

We model the advisor-client relationship as a cheap-talk game between two players with a potential conflict of interests between them (see the seminal paper in this area by Crawford and Sobel, 1982). In each period, an Advisor (she) and a Client (he) are matched with each other. The Client needs to make a decision that will affect both his own and the Advisor's payoffs. In particular, the Client needs to choose an action, $a \in \{1, 2, ..., 10\}$ that should match the realized state, $s \in \{1, 2, ..., 10\}$ (uniformly distributed). When choosing his action, the Client does not know s, but he observes a message, $m \in \{1, 2, ..., 10\}$ from the Advisor who exactly knows s. This message might or might not be truthful, and it is never payoff-relevant.

Payoffs in this game only depend on a, the chosen action by the Client. The Client's payoff function is concave with achieving highest payoff when the chosen action is exactly the realized state: $\Pi^C = 100 - (s - a)^2$.[5] There are two types of Advisor: one with aligned and one with misaligned preferences. Advisors with aligned preferences ("aligned Advisors") have the same preferences as the Client: $\Pi^A_u = 100 - (s - a)^2$, whereas the misaligned types earn the highest payoff when the Client chooses 10: $\Pi^A_b = 100 - (10 - a)^2$, irrespective of the realized state. Each type of Advisor is equally likely: with 50% probability the Advisor's preferences are aligned, with 50% probability they are misaligned. The Advisor knows her own type, whereas the Client only knows the type in certain treatments of the experiment (see Section 3.2).

The sequence of events in a period is the following:

1. The Advisor observes her type and s, and sends a message m to the Client (and in some treatments decides whether the Client observes her type).

2. The Client observes the message m, and in some treatments the type of the Advisor, and then chooses an action a.

[5] A similar quadratic loss function was also applied in Crawford and Sobel (1982), and in many subsequent research on strategic communication (see e.g., Gesche, 2021 or Barron and Fries, 2024). We add a constant so that in the experiment players do not incur high losses.

3. Payoffs realize, and each player receives feedback about their own payoffs and the Client's action a. (s is not disclosed to the Client.)

We repeat the game for several periods in three blocks. Participants keep their role throughout the entire experiment. The Advisor's type and the realized state s are randomly and independently drawn in each period. We implement a block termination procedure in each block. The continuation probability in a block is 90% in each period. A random draw after each period determines whether the block continues or stops, and the experiment either moves to the next block or ends. However, in order to have enough observations in each block, we let participants play 10 periods without telling them the random number. We reveal all the random draws from the first 10 periods after period 10, and the block stops if any of these random draws results in termination. After that, the random draw is revealed after each period. We carefully explain this procedure to the participants (see Appendix C.1 for instructions).

3.2 Treatments

We implement a 2×3 between-subject experimental design where we vary the interaction length between Advisor and Client, and the disclosure protocol in place, resulting in six treatments in total. On the interaction length dimension, we consider a Short-term treatment and a Partner treatment:

- Under **Short-term treatment**, Advisors and Clients are re-matched after each period. They are told that they are never matched with the same person in two subsequent periods, and that they would never receive any personal identification of the partner. Essentially we implement an imperfect stranger matching in this treatment.

- Under **Partner treatment**, pairs are matched for a block. Here participants are told that they are never matched with the same partner for two consecutive blocks.

Even though it has less influence on the matching, we keep the block structure also in the Short-term treatment to have the procedure and the instructions as comparable as possible across the treatments. We re-matched participants in a matching group of six participants (three advisors and three clients).

When considering the disclosure protocol, we implement three different protocols regarding the disclosure of the Advisor's type:

- Under **No disclosure**, the Client is not informed about the type of the Advisor, and is reminded that each type is equally likely together with the payoffs of the different types.

- Under **Mandatory disclosure**, the Client is informed in each period about the Advisor's type, and is reminded of the payoff of that type.

- Under **Voluntary disclosure**, the Advisor has the possibility to truthfully disclose her type. When making her decision about the message, she also decides whether to tell her type to the Client. In case she decides to

Table 1: Treatment summary

	No disclosure	Mandatory disclosure	Voluntary disclosure
Short-term matching	SN	SM	SV
Partner matching	PN	PM	PV

tell the Client, the Client learns the true type of the Advisor. In case the Advisor does not disclose her type, the Client learns that the Advisor decided not to disclose and he is reminded about the possible types.[6]

Combining the two interaction lengths and the three disclosure procedures leads to six distinct treatments summarized in Table 1.

3.3 Theoretical predictions and hypotheses

We start by looking at the stage game with the standard rationality assumption. While there are multiple equilibria, we will focus in this section on the ones that transmit the most information. We relegate the proofs of the lemmas, as well as examples of further equilibria to Appendix A. After discussing the Nash equilibria, we turn to behavioral explanations as well based on previous literature and we formulate our hypotheses. Given that we have rich data, we decided to formulate preregistered hypotheses only on welfare (payoffs), but not on other aspects of behavior that drive the results on welfare. We will analyze the decisions as exploratory analyses.

Under *No disclosure* it can be shown that the most informative equilibrium is a semi-separating one characterized by the following strategies:

Lemma 1 (Stage game under no disclosure) *The most informative Perfect Bayesian Nash equilibrium is characterized by the following strategies and beliefs:*

- *An aligned Advisor chooses* $m = s$ *when* $s \leq 6$*, and randomizes between messages* $m \in \{7, 8, 9, 10\}$ *for* $s \in \{7, 8, 9, 10\}$.
- *A misaligned Advisor always chooses a message randomly from* $m \in \{7, 8, 9, 10\}$ *irrespective of* s.
- *The Client chooses* $a = m$ *when* $m \leq 6$*, and holds the belief that the Advisor is aligned with probability 1.*
- *The Client chooses* $a = 6$ *when* $m > 6$*, and holds the belief that the Advisor is aligned with probability* $2/7$.

Under *Mandatory disclosure* the most informative equilibrium is when the aligned Advisor tells the true state in each state. However, the misaligned Advisor cannot transmit any information in this situation.

Lemma 2 (Stage game under mandatory disclosure) *The most informative Perfect Bayesian Nash equilibrium is characterized by the following strategies and beliefs:*

[6] Note that randomly drawing the Advisor's type in every period makes the disclosure decision relevant in each period in both treatments. While there are some arguments why in real world COI would not change that easily over time, there are also situation where it is realistic to assume that the Advisor's interests change relatively quickly (e.g., a financial advisor might have interests in different stocks in different periods making them more or less aligned with the Client's preferences as time proceeds).

- *An aligned Advisor chooses $m = s$ for all s, whereas a misaligned Advisor always chooses a message randomly for all s.*
- *The Client chooses $a = m$ when the Advisor is aligned, and $a = 5$ (or $a = 6$) when the Advisor is misaligned.*

Under *Voluntary disclosure* the equilibrium outcome would be characterized by either the No disclosure case or the Mandatory disclosure case, as if one type of Advisor optimally discloses her type in equilibrium, then the other type is also known. When neither type finds it beneficial to disclose the conflict, then the resulting outcome can be described by the No disclosure case (see also Li and Madarász, 2008). Given that there are many equilibria in both scenarios ranging from no information transmission at all to partial information transmission, it is possible to construct equilibria where either type of disclosure decisions are sustainable even though equilibria with no disclosure might be intuitively less likely (e.g., babbling in case the types would be revealed, and partial information transmission under no disclosure). Combining the two previous disclosure protocols, we find that the most informative equilibrium involves disclosure.

Lemma 3 (Stage game under voluntary disclosure) *The most informative Perfect Bayesian Nash equilibrium is characterized by the following strategies and beliefs:*

- *Both types of Advisor disclose her type.*
- *In the subgame after disclosure, players follow the strategies from Lemma 2.*
- *In the subgame after no disclosure, players follow the strategies from Lemma 1.*

The proof of Lemma 3 can be easily seen by observing that the misaligned Advisor is indifferent between disclosing or not, whereas the aligned Advisor strictly prefers disclosure, as then she can transfer more information to the Client. In Lemma 1 and 2 we have established the equilibria for the subgames.

When introducing indefinite repetition to the stage game, then there are even more possibilities in equilibrium. Provided that players are sufficiently patient, they can achieve higher payoffs than in the stage game (see the folk theorem in game theory) where the benchmark under disclosure is the expected payoff ex-ante before drawing the Advisor's type. Given that even in the most informative equilibria there is room for payoff improvement when the Advisor has misaligned preferences, there are equilibria in the repeated setting that can maintain higher payoffs compared to the equivalent stage game. This already suggests that even without behavioral economic mechanisms, repeated play can be expected to raise payoffs.

Given that under the rational assumptions multiple equilibria exist, we do not only look at this assumptions, but also at behavioral explanations to derive our hypothesis. We start with the repeated interactions. Besides the positive prediction based on folk's theorem, repeated interactions with the same partner allow players to learn about the partner's behavior. In particular, Clients can learn about the Advisors' messaging strategy, and Advisors can learn about Clients' reaction to their messages. Also, Advisors might want to build a reputation, and transmit more information even when their interests are not aligned with the client, or Clients might want to reward misaligned

Advisors by choosing an action that harms themselves a bit, but helps the Advisor more.[7] We expect therefore that under repeated interactions Advisors tell more often the truth, and that Clients discount messages less often or less intensively. These forces all point in the same direction leading to our first hypothesis about the effect of introducing long-term partnership on welfare:

Hypothesis 1 (Effect of interaction length) *Keeping the disclosure protocol fixed, average payoffs of both Advisors and Clients will be higher in the Partner than in Short-term treatment.*

Turning our attention to the differences across the disclosure protocols, we distinguish between Short-term and Partner interactions. Looking at the stage game, we have already established that there are multiple equilibria under all three disclosure protocols. When considering the most informative equilibria, based on the rationality assumption we would expect that Voluntary disclosure and Mandatory disclosure result in the same outcome, and they both outperform No disclosure. However, there are already several theoretical and experimental papers that compare No disclosure to Mandatory disclosure in different settings. Most of these papers show that Mandatory disclosure can hurt rather than help.[8] Based in the previous literature showing a potential harmful effect of Mandatory disclosure we hypothesize that disclosure can backfire in Short-term relations for both Advisors and Clients. This because misaligned Advisors feel licensed to lie and exaggerate more in their advice, whereas Clients do not discount the advise enough.[9] When turning to Voluntary disclosure there is not much evidence that we can rely on to formulate our hypothesis. Sah and Loewenstein (2014) finds that voluntary disclosure helps clients, but an important difference from our study is that advisors can decide whether to accept or reject the COI in the first place. Therefore, the share of misaligned advisors is not constant between treatments, making comparisons more difficult. Given this lack of evidence and that theoretically the voluntary disclosure should fall in between the two exogenous disclosure treatments, we expect that it is not different from the empirical best of these two, i.e., No disclosure. Summarizing these arguments, we can state our second hypothesis.

Hypothesis 2 (Effect of disclosure protocols in the Short-term treatments)

a *Average payoffs of both Advisors and Clients will be lower in the Mandatory than in the No disclosure treatment.*

b *Average payoffs of both Advisors and Clients will not be different in the Voluntary compared to the No disclosure treatment.*

[7] There is support for the positive effect of repetition both in the theoretical (Sobel, 1985; Stocken, 2000) and the experimental (Koch and Schmidt, 2010) cheap talk literature.

[8] Li and Madarász (2008) theoretically show that Mandatory disclosure can backfire because advisors exaggerate their bias more (due to moral licensing), and clients discount too much. In experiments Cain et al. (2005), Cain et al. (2011) show that clients do not discount enough to account for the higher advice from the advisor resulting in lower payoffs for clients if they are matched with a conflicted advisor. There are also papers that show that Mandatory disclosure neither helps nor hurts (Ismayilov and Potters, 2013 or Church and Kuang, 2009). Note that the research mentioned in this footnote was conducted in settings where aligned and misaligned advisors were separated between treatments unlike in our setting.

[9] One could argue that if clients are hurt, advisors must benefit, but this is not necessarily the case. For example, Cain et al. (2005) only finds an effect on clients but not on advisors.

c Average payoffs of both Advisors and Clients will be higher in the Voluntary than in the Mandatory disclosure treatment.

When we have a repeated interaction, we expect that the learning and reputation building help by utilizing the additional information that any type of disclosure provides participants. The fear of potential punishment will compete with a potential license to lie by misaligned Advisors resulting in less biased advise. As an exchange, Clients learn to use the disclosed information and interpret the message better than under no disclosure. Furthermore, as Voluntary disclosure can serve as a signaling device to build trust, we expect that outcomes improve even further under Voluntary compared to Mandatory disclosure. When an Advisor voluntary makes her vulnerable by disclosing their misaligned preferences, that can serve as a signal of trust for the Client. The Client might expect the Advisor then to give less biased advise, and therefore follow it more often. That disclosure helps under partnership was also shown by Koch and Schmidt (2010) in the setting where the authors only focused on misaligned advisors. This leads us to our third and last hypothesis.

Hypothesis 3 (Effect of disclosure protocols in the Partner treatments)

a Average payoffs of both Advisors and Clients will be higher in the Mandatory than in the No disclosure treatment.

b Average payoffs of both Advisors and Clients will be higher in the Voluntary than in the No disclosure treatment.

c Average payoffs of both Advisors and Clients will be higher in the Voluntary than in the Mandatory disclosure treatment.

3.4 Further tasks

In order to better understand individual decisions, we collect ex-post (unincentivized) beliefs and let participants play an incentivized number-assigning task to assess their individual lying preferences. Advisors are asked about their beliefs regarding Clients following specific messages. Under No disclosure, we tell nothing about the Advisor. For the other two treatments, we collect Advisor beliefs assuming that it is either known that the Advisor is misaligned, or in the Voluntary treatments assuming that the Advisor did not share her type.

Clients were asked about their beliefs about the share of Advisors who sent a message equal to the true state for certain messages. Again, we asked different types of Advisors in the different treatments. Under No disclosure, we asked about Advisors in general, whereas in the other two disclosure treatments, we asked about a misaligned Advisor. Furthermore, under Voluntary disclosure, we also measured Clients' beliefs regarding the disclosure decision of a misaligned and an aligned Advisor.

After the belief-elicitation, all participants answered the same number assignment task. They were assigned a number (an integer from 1 to 6), and they had to claim the assigned number. However, they were allowed to claim any number irrespective of the actually assigned number. The payment for this task depended only on the claimed number, not on the assigned one. In particular, the payment is determined by the following formula: 100 ECU +

20 * claimed number. We applied a strategy method here and asked participants to state their claimed number for each possible assigned number. Earnings were based on a uniform random draw from the assigned numbers. The corresponding claimed number was relevant for earnings.

After this task, participants answered a short questionnaire in which we assessed general trust[10], and collected gender, age and field of study. Instructions for this second part are in Appendix C.3.

3.5 Procedures

The experiment was programmed in PHP, and the sessions were run in July, October and December 2024 in the MELESSA lab at LMU Munich. The experiment was preregistered prior running the first session.[11] We have collected 12 independent matching groups for each of the six treatments. In total 456 participants participated in 34 sessions. None of the participants in our final sample participated more than once.[12] Sessions lasted about 1.5 hour, and the average payoff was €27.14 including a €10 show up fee. We collected data from 3 blocks for most matching groups. Two groups (in SV) managed to finish only two blocks in the time allocated for the session, and four groups (one in SV, two in PN, and one in PM) finished four blocks of the game.

Given that the experiment contained substantial random draws (realized states and type of Advisor), we have run the randomization before the experiment, and fixed it for each matching group. This means that each matching group in one treatment faced a unique sequence of random draws, but these sequences were mirrored in all treatments allowing us to make better inferences across treatments, as the random draws are comparable. The randomization is included in the preregistration, and contained the random draws of the preferences and the state of the world (unique per matching group), as well as the random numbers determining the length of the sessions (same draw for all the sessions).

Furthermore, in order to avoid order effects, we randomized on the individual level the order of information on the decision screens.[13] For the Advisor we randomized the order of the information about her preferences and the realized state. In the Voluntary treatments, the order of information also matched the order of the two decisions (disclosure and message decisions). For the Client, we randomized the order of the Advisor's message and her preferences if known (if preferences were unknown, the Client still received information about potential preferences). Experimental instructions and screenshots for treatment Partner Voluntary can be found in Appendix C.

[10] De Meza et al. (2010) show that trust can be an important factor in persuasion situations.

[11] See preregistration at OSF link.

[12] We rerun 4 matching groups because it turned out that due to an administrative issue, 3 participants participated multiple times. We have dropped the matching groups in which they have participated for the second or third time, and collected new groups instead of those. The results do not depend on these changes.

[13] Saccardo and Serra-Garcia (2023) shows that the order in which information is presented to advisors might matter for their recommendation. In their paper, advisors who saw that they had a conflict of interests first, recommended more often the product that was worse for the client than advisors who saw the quality of the product first. This quality can be analogous to the state of the world in our experiment.

4 Results

We first present Advisors' communication behavior and Clients' responses to these messages to investigate how the different treatments affect behavior, as behavioral change can result in welfare change. These analyses are all exploratory, they were not part of the preregistration. After understanding these behavioral forces, we investigate whether differences lead to different welfare outcomes, corresponding to our pre-registered hypotheses where we were interested in the overall effects of relationship length and disclosure protocol separately for Advisors and Clients. Following our preregistered analysis plan, statistical comparisons for the pre-registered hypotheses use Mann-Whitney rank-sum tests with the matching group as the unit of observation. Directional hypotheses are evaluated using one-tailed tests. The unit of analysis includes behavior in the first three blocks (as very few matching groups reached block 4).

4.1 Behavioral analysis

In this section, we focus on how disclosure and relationship length affect disclosure decisions, exaggeration, trust in advice, and ultimately information transmission.

4.1.1 Advisors' behavior

Advisors' voluntary disclosure decision. When preferences are aligned, Advisors are disclosing this 94% of the time in short-term interactions. This rate is slightly lower (86%) in partner interactions. This pattern is intuitive: when interests are aligned, disclosure is costless and reassures Clients about the Advisor's credibility, while repeated interaction may itself foster trust and reduce the need for explicit reassurance. Furthermore, when aligned Advisors always disclose in a repeated interaction, then no disclosure immediately signals misaligned preferences. When misalignment is present, Advisors are disclosing this much less frequently in both treatments: 66% of the time in short-term interactions and 64% in partner ones.

Table 2 formally examines whether disclosure decisions depend on advisors' incentives, lying preferences[14], and interaction length. We find that advisors are significantly less likely to disclose when a conflict of interests is present ($\beta = -0.289$, $p < 0.01$). Advisors with high lying aversion disclose less frequently, although this effect is only marginally significant ($\beta = -0.098$, $p = 0.074$). Partner matching does not significantly affect disclosure behavior. We do not detect significant interaction effects, suggesting that the relationship between preferences, lying aversion, and disclosure is stable across interaction lengths.

Result 1 *The main driver of voluntary disclosure of COI is the Advisor's preference: having misaligned preferences significantly reduces the probability of disclosure by 29 percentage points. This pattern is stable across short-term and partner interactions.*

These findings suggest that disclosure decisions are primarily driven by strategic considerations rather than relationship length. Advisors disclose substantially less often when they have a conflict of interest, consistent with

[14]We classify a participant as having low lying aversion if they lie at least once for personal gain and never in a way that harms them. Otherwise, we classify them as having high lying aversion.

Table 2: Determinants of Advisors' Disclosure Decisions

	Disclosure
Misaligned preferences	-0.289***
	(0.055)
High lying aversion	-0.098*
	(0.055)
Partner interaction	-0.077
	(0.058)
Misaligned × High lying aversion	0.035
	(0.057)
Misaligned × Partner	0.050
	(0.074)
High lying aversion × Partner	0.014
	(0.074)
Constant	0.983***
	(0.018)
Observations	2628

Notes: Linear probability model. The dependent variable equals one if the advisor disclosed their type. Robust standard errors clustered at the matching-group level are reported in parentheses. Misaligned preferences equals one if the advisor had a conflict of interest. High lying aversion equals one for advisors classified as having high lying aversion. Partner interaction equals one in repeated interactions. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

incentives to withhold unfavorable information. The absence of significant interaction effects indicates that long-term relationships do not substantially alter these disclosure incentives. Although only marginally significant, the negative association between lying aversion and disclosure may reflect differences in how advisors perceive moral costs in this multi-dimensional communication setting. Advisors with stronger aversion to lying may prefer to avoid active misrepresentation in their messages, while perceiving nondisclosure as less morally problematic. Moreover, if receivers discount advice more strongly when misaligned preferences are disclosed, nondisclosure may allow advisors to maintain influence without engaging in explicit deception.

Advisors' messaging behavior. For a more systematic analysis of Advisors' messaging strategy, we assign messages to three categories: i) those that "exaggerate", if the message is larger than the state, ii) those that tell the "truth", if the message is equal to the state, and iii) those that "underreport", if the message is smaller than the state. Figure 1 presents the distribution across treatments of these three types of messaging strategies employed by Advisors when their preferences are *aligned* with the Clients. In this case, truthful messaging is the dominant strategy. The high baseline rate of truthful messages is therefore consistent with equilibrium predictions. Small deviations from truth when the preferences are known (right panel) likely reflect noise or mistakes rather than strategic misrepresentation. When preferences are not disclosed, under-reporting can be strategic: Clients might discount higher messages more, therefore in high states an aligned Advisor might not want to disclose the true state to induce less discounting. In addition, we find significantly (though modestly) larger rates of truth-telling from aligned Advisors under disclosure, particularly when this is voluntary (see Table 3 in Appendix for regression results). This suggests that transparency could reduce information distortion even when incentives are aligned.

Result 2 *In absence of conflict of interests, Advisors almost always report the state truthfully.*

Figure 2 presents the distribution across treatments of the three types of messaging strategies employed by Advisors when their preferences are *misaligned* and they have conflicting interests with the Clients. The first observation is that by far, the most frequent type of message from misaligned Advisors is exaggeration both when the COI is disclosed and when this is not the case. This happens in all treatments except for PV, when the COI is disclosed, where truth-telling is more frequent. Under-reporting is stable and relatively infrequent in all treatments, so any changes in exaggeration rates correspond to similarly sized changes in truth-telling rather than under-reporting rates. This pattern is consistent with strategic communication under COI: exaggeration is expected to shift the Client's action toward the Advisor's preferred outcome. The prevalence of exaggeration therefore reflects incentive-driven distortion.

When investigating the effect of interaction length, we find that overall, partner settings reduce the likelihood of exaggeration by about 6 percentage points ($p = 0.019$, see Table 4 in Appendix for full regression results). This is a more direct evidence of the hypothesized mechanism behind the benefits of partner settings: repeated interaction can discipline exaggeration by introducing reputational incentives. Advisors who exaggerate excessively risk losing credibility and influence in subsequent rounds, while more informative communication can sustain trust and responsiveness over time.

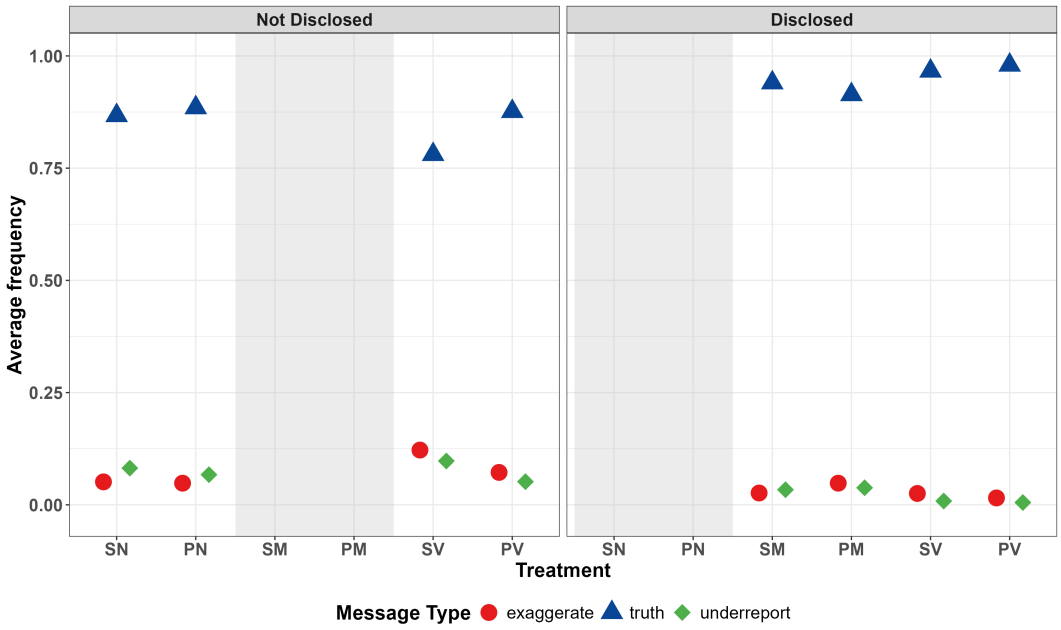


Figure 1: Aligned Advisors' Message Type Frequency Distribution Across Treatments

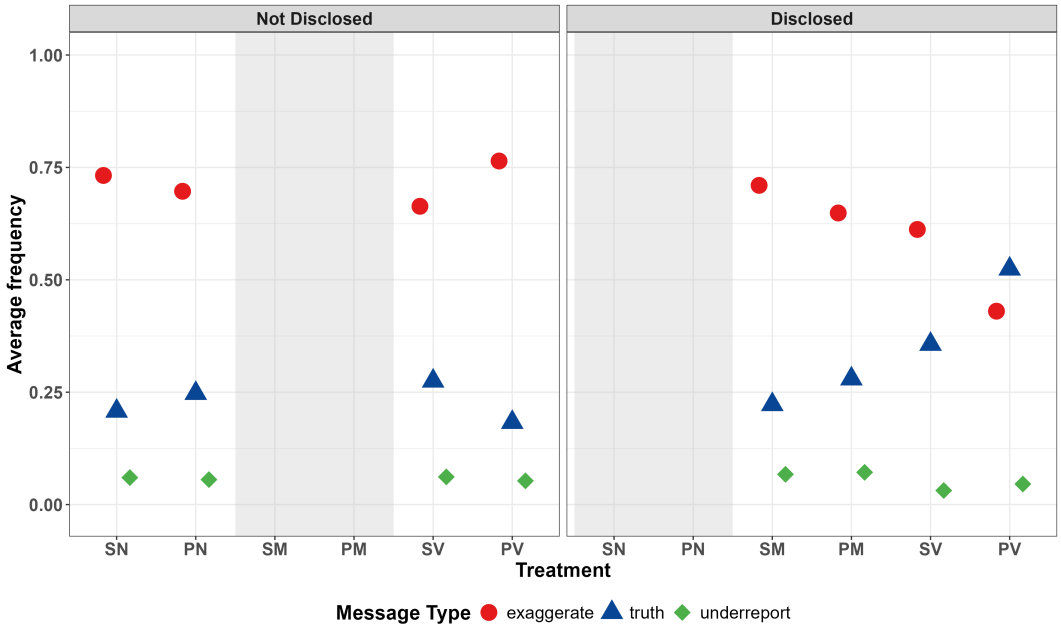


Figure 2: Misaligned Advisors' Message Type Frequency Distribution Across Treatments

As Figure 2 shows, this effect is mostly driven by the voluntary disclosure setting, when the COI is disclosed. Indeed, when we focus on those cases, we find that the exaggeration rate in the PV treatment is about 18 percentage points lower than that in SV ($p < 0.001$, see Table 4 in Appendix for full regression results). The small increase in exaggeration in PV compared to SV when the bias is not disclosed is not significantly different. The particularly large reduction in exaggeration under partner matching after voluntary disclosure suggests a complementary effect between relational incentives and transparency. When disclosure is voluntary, choosing transparency can signal credibility, while repeated interaction increases the value of maintaining trust. Together, these forces make exaggeration less attractive and truthful communication more prevalent.

Contrary to concerns that mandatory disclosure may induce strategic exaggeration to offset anticipated skepticism, we find no evidence that mandatory disclosure increases exaggeration in neither partner nor short-term interactions.

Result 3 *When Advisors have conflicting interests, exaggeration is the dominant communication strategy across treatments. Partner matching reduces exaggeration modestly overall. This reduction is driven by the voluntary disclosure setting: when COI is voluntarily disclosed, exaggeration falls substantially under partner matching and truthful communication becomes the majority strategy. In contrast, mandatory disclosure does not increase exaggeration.*

4.1.2 Clients' behavior

Clients' response to messages. We assign Clients' responses to Advisors' messages into three categories: i) those that "discount", if the action is lower than the message, ii) those that "follow", if the action is equal to the message, and iii) those that "reward", if the action is larger than the message. Figure 3 presents the distribution of each action type frequency across treatments for cases where the Advisors' preferences remained undisclosed, either because this was forced or because the Advisor voluntarily chose so. Tables 5 and 6 in the Appendix present the full regression analyses of this behavior.

Overall, we see that voluntary non-disclosure leads to more skeptical responses from Clients, with follow rates being significantly lower in these treatments than when no-disclosure is forced ($\beta = -0.15$, $p < 0.001$). The figure also reveals a positive effect of partner interactions on following rates. However, this effect is only marginally significant ($\beta = 0.10$, $p = 0.088$). In the forced no-disclosure setting, this increase leads to a small and insignificant decrease in discounting behavior due to partner interaction, but to a significant decrease in rewarding behavior ($\beta = -0.05$, $p = 0.017$). In the voluntary no-disclosure, Clients are slightly more likely to reward non disclosing partners ($\beta = 0.06$, $p = 0.082$), and much less likely to discount their message ($\beta = -0.15$, $p = 0.012$).

Result 4 *When the Advisors' preferences are not disclosed, Clients respond more skeptically in voluntary disclosure environments than when non-disclosure is imposed exogenously, following advice less frequently. Partner interaction mitigates skepticism, particularly in the voluntary setting, leading to reduced discounting.*

Figure 4 presents the distribution of each action type frequency across treatments for cases where the Advisors' preferences were disclosed, either because this was mandated or because the Advisor voluntarily chose to do so.

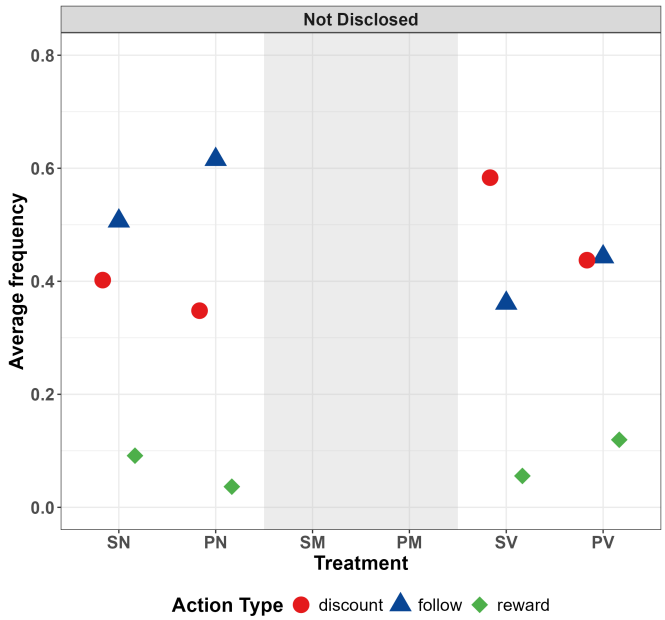


Figure 3: Clients' Action Type Frequency Distribution Across Treatments when Preferences are Not Disclosed

When the Advisor has aligned preferences, Clients almost always follow their message (as they should). When the Advisor has misaligned preferences, and the disclosure is mandatory, Clients' response to messages is not affected by interaction length. However, when disclosure is a choice, being in a partner interaction compared to a short-term one, significantly increases trust in messages: follow rates are 14 percentage points higher ($p = 0.006$), discount rates are 22 percentage points lower ($p = 0.001$), and reward rates are 8 percentage points higher ($p = 0.028$).

Result 5 *When Clients find out that the Advisor's preferences are aligned, Clients almost always follow recommendations, irrespective of the disclosure protocol. When Advisors have misaligned preferences and disclosure is mandatory, Client responses do not vary with interaction length. In contrast, when disclosure results from the Advisor's choice, partner interaction substantially increases trust.*

The increase in rewarding behavior in partner interactions is consistent with Clients responding to long-term incentives: rewarding may help sustain credibility and promote future transparency and truthful communication. Moreover, this behavior can be efficiency-enhancing given the quadratic payoff structure. Small upward deviations from the Client's preferred action reduce the Client's payoff only modestly while might increase the misaligned Advisor's payoff more substantially. Consistent with this asymmetry, when Advisors voluntarily disclose their misaligned preferences in the partner setting, pairs earn significantly more when Clients reward rather than simply follow the message (two-sided MW test, $p = 0.001$).

Clients' action accuracy. While the previous subsections showed how voluntary disclosure in partner settings increases truthful information transmission from Advisors, the analysis of Clients' response to messages only indirectly suggests that this information is also received. To get a more accurate picture of how much partnership and

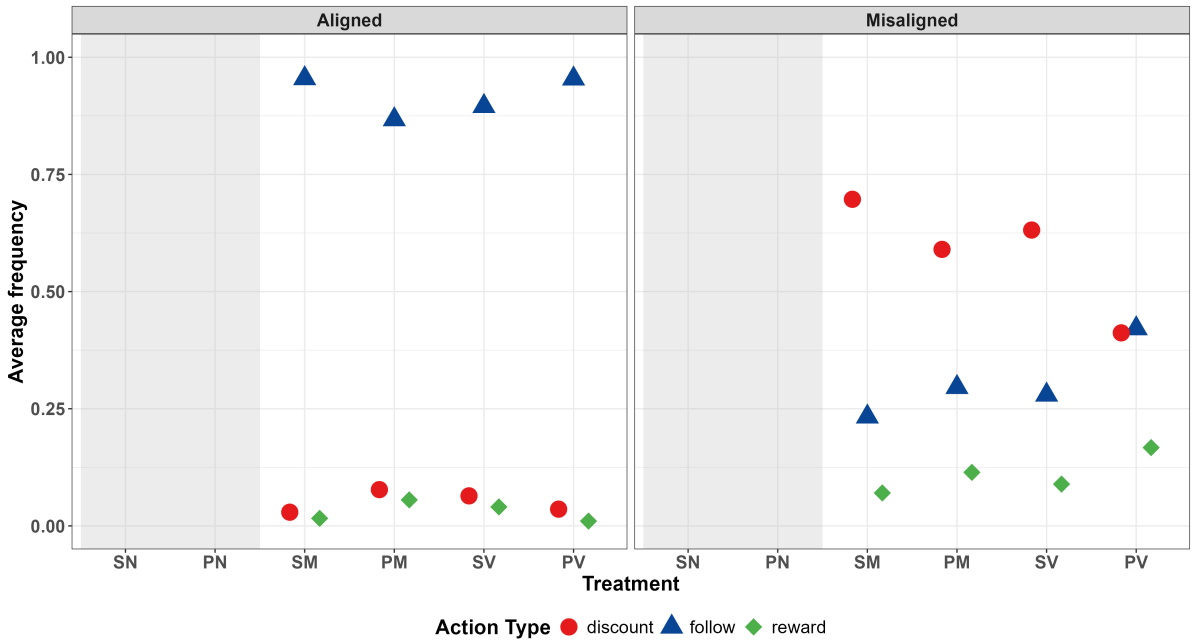


Figure 4: Clients' Action Type Frequency Distribution Across Treatments when Preferences are Disclosed

voluntary disclosure help information reception, we analyze the accuracy of Clients' actions, i.e. how often their action matched the randomly drawn state.

Figure 5 presents accuracy rates for Clients' actions, across the short-term and partner settings, depending on whether the preferences were disclosed or not, and whether this (non-)disclosure was voluntary or exogenously imposed.

First, unsurprisingly, accuracy rates are considerably higher when there is no misalignment in preferences and this accuracy increases when the preferences are disclosed exogenously or voluntarily (see Table 8 in Appendix). Next, we find an overall weakly significant effect of partner interaction, which increases the action accuracy by 4 percentage points ($p = 0.083$, Table 9 in Appendix). This increase is larger when we focus on the voluntary disclosure cases where misalignment was disclosed ($\beta = 0.12$, $p = 0.013$, Table 9 in Appendix).

Turning to the effect of voluntary disclosure, we find a significantly positive effect of disclosure on the accuracy rate which increases with 12.6 percentage points compared to mandatory disclosure when preferences are disclosed ($p < 0.001$, Table 10 in Appendix). When preferences remain undisclosed, the voluntary protocol leads to a decrease in accuracy rates of 15 percentage points ($p < 0.001$, Table 10 in Appendix). While there are considerably fewer instances of voluntary non-disclosure (595 vs. 2736), this is in line with Clients being more likely to punish voluntary non-disclosure. In these cases of non-disclosure, partner interaction has a stronger positive effect ($\beta = 0.08$, $p = 0.020$, Table 10 in Appendix).

When we pool over all (non-)disclosure decisions in the voluntary regime and compare it with the exogenously imposed no-disclosure protocol, we find a significant increase in accuracy rates of 14 percentage points ($p < 0.001$, Table 11 in Appendix). Mandatory disclosure also leads to a significant increase in accuracy rates of 10 percentage points compared to no-disclosure ($p < 0.001$, Table 11 in Appendix). The difference between the voluntary and the

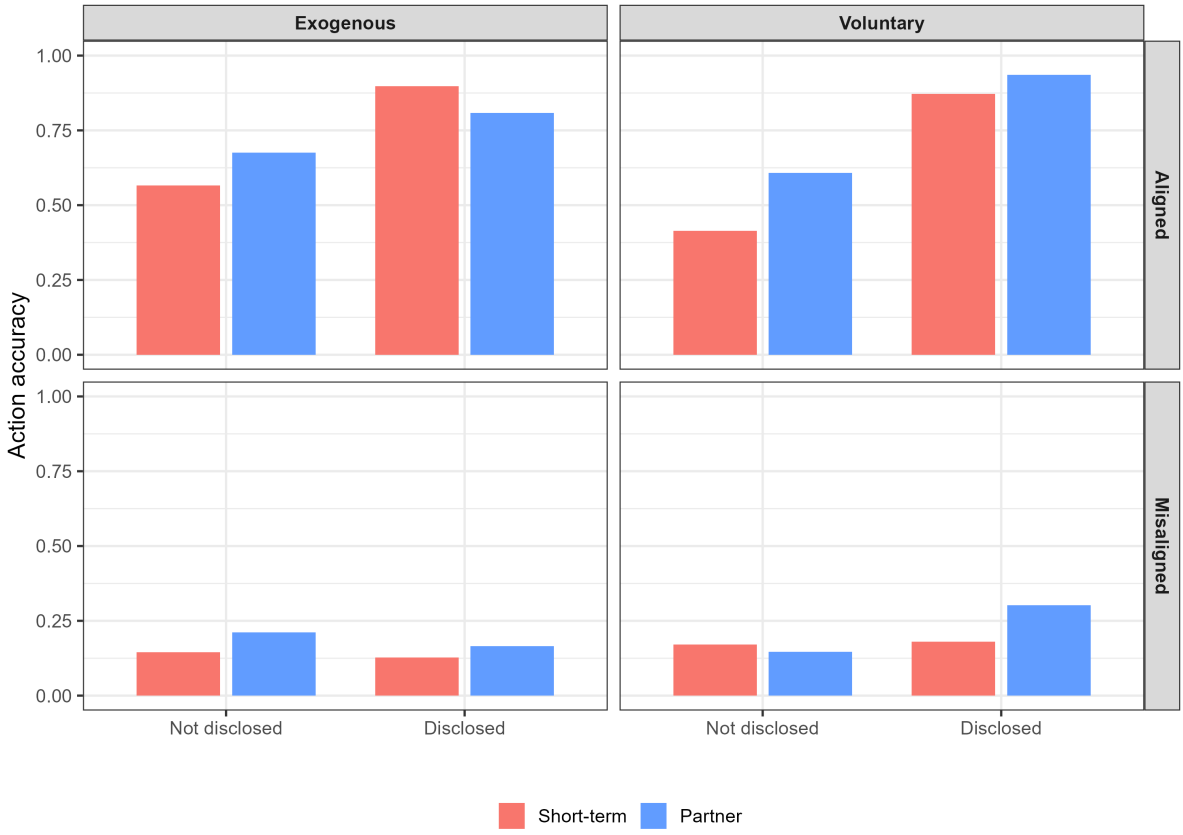


Figure 5: Clients' Action Accuracy Rates Across Treatments

mandatory protocols is not significant. This result suggests that disclosure helps identifying aligned Advisors.

Result 6 *Decision accuracy improves when preferences are disclosed and declines when preferences remain undisclosed. Partner interaction modestly improves accuracy overall and substantially improves it when COI is voluntarily disclosed. Both voluntary and mandatory disclosure improve decision accuracy relative to no disclosure, while voluntary nondisclosure is interpreted negatively by clients.*

4.2 Welfare analysis

In the previous subsection we have seen that both Advisors (mainly with misaligned preferences) and Clients react to treatment differences. Misaligned Advisors exaggerate in their messages, but they do so in a lesser extend in long-term relations. This results is mainly driven by the voluntary disclosure, where after disclosing misaligned preferences, majority of Advisors message in partner settings is truthful. Given that misaligned Advisors disclose their type in about 65% of the cases, the result is not driven by only a few data-points.

When turning to the Clients, we find that partner interactions decrease discounts in advise, and this especially the case when facing a misaligned Advisor who disclosed her type. When the Advisor chose not to disclose their preferences, that is interpreted as misaligned by Clients, and are punished compared to when preferences were forced to remain hidden. All in all, taking all these behavioral effects together, voluntary disclosure increases trust and

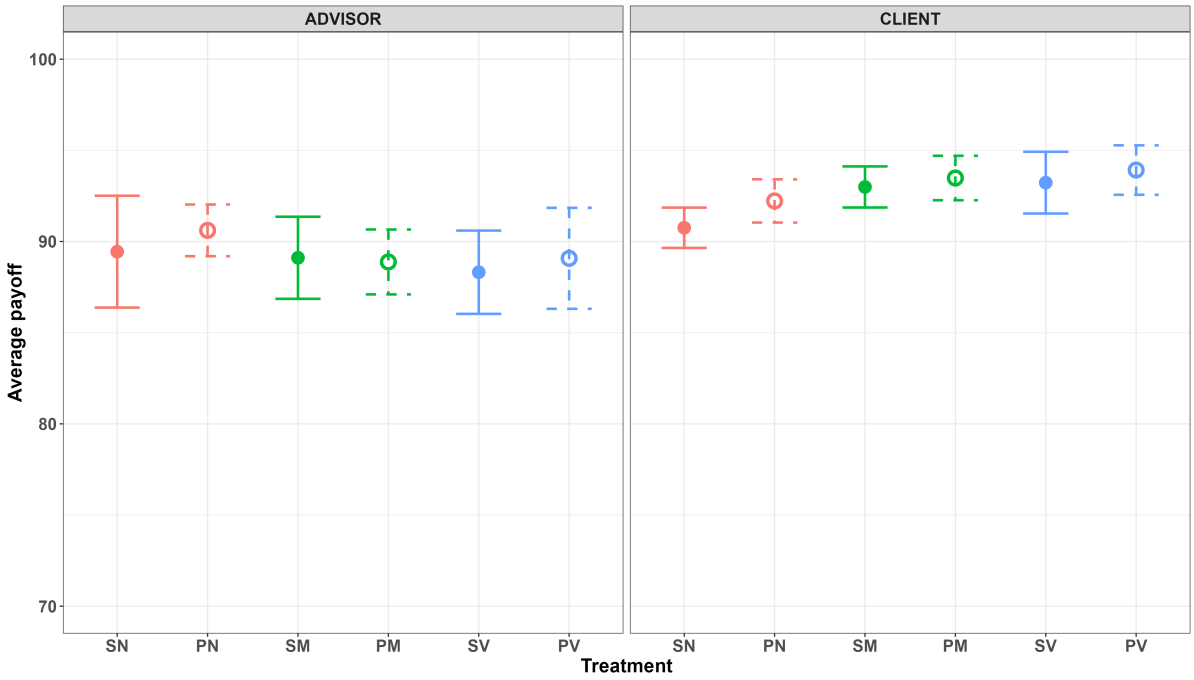


Figure 6: Overall payoff of Advisors and Clients across treatments

trustworthiness, particularly in a partner setting. These findings also affect decision accuracy. In this subsection we look at how the differences in behavior translate into realized welfare, i.e. we are turning to our pre-registered hypotheses.

Figure 6 presents the distribution of average payoffs for the two players across treatments. The first set of hypotheses relates to the effect of relationship length on players' average payoff. We find no significant differences between Advisors' average payoffs across the S- and P- treatments, holding constant the disclosure protocol (SN vs. PN, MW test, $p = 0.534$; SM vs. PM, MW test, $p = 0.622$; SV vs. PV, MW test, $p = 0.378$). Clients also do not benefit from partner interaction when mandatory or voluntary disclosure is in place (SM vs. PM, MW test, $p = 0.315$; SV vs. PV, MW test, $p = 0.443$), but they do reach significantly higher average payoffs in the partner setting compared to the short-term one when no disclosure is possible (SN vs. PN, MW test, $p = 0.050$).

Result 7 *Partner matching does not have an effect on players' average welfare, except for Clients when no disclosure possibility is available.*

We now move on to the effect of the disclosure protocol within the short-term interaction setting. Advisors' payoffs in short-term interactions are not significantly different across the disclosure protocols (SN vs. SM, MW test, $p = 0.189$; SM vs. SV, MW test, $p = 0.378$; SN vs. SV, MW test, $p = 0.204$). However, clients in short-term interactions significantly benefit from voluntary disclosure compared to no disclosure (SN vs. SV, MW test, $p = 0.019$), but this benefit is not significantly larger when compared to mandatory disclosure (SM vs. SV, MW test, $p = 0.221$). Clients are also benefiting from mandatory disclosure compared to no disclosure in the short-term setting we implement (SN vs. SM, MW test, $p = 0.008$).

Result 8 *In short-term interactions, Advisors' average payoffs are not affected by either mandatory or voluntary disclosure. Clients, however, are benefiting both from mandatory and voluntary disclosure compared to no disclosure.*

This pattern is different than what we hypothesized for the short-term interactions. We expected that mandatory disclosure will be worse than no disclosure which will be just as good as voluntary. The theoretical analysis can explain the reversal nevertheless, as the voluntary disclosure protocol should simply approximate the better of the two forced disclosure settings, which turned out to be the mandatory disclosure one in our experiment. In Section 4.1.1 we have seen that misaligned Advisors do not feel licensed to lie in our experiment. The exaggeration rates between the no disclosure and mandatory disclosure settings are similar. This can explain why mandatory disclosure does not harm in our experiment.

In the repeated interaction setting, Advisors are again not benefiting (as we would have expected) nor being hurt by any disclosure protocol (PN vs. PM, MW test, $p = 0.928$; PM vs. PV, MW test, $p = 0.335$; PN vs. PV, MW test, $p = 0.665$). Clients again benefit, but this time only from voluntary disclosure (PN vs. PM, MW test, $p = 0.146$; PM vs. PV, MW test, $p = 0.292$; PN vs. PV, MW test, $p = 0.039$)

Result 9 *In partner interactions, Advisors' average payoffs are again not affected by either mandatory or voluntary disclosure. Clients, are benefiting only from voluntary disclosure compared to no disclosure.*

Overall, we find that Clients systematically benefit from voluntary disclosure, and sometimes by mandatory disclosure as well, while Advisors are not hurt by either type of disclosure. While these results provide a useful benchmark, the role of disclosure is fundamentally tied to conflict of interests. When Advisors' incentives are aligned with those of the Clients, truthful communication is incentive compatible, and disclosure primarily serves a reassurance function for the Clients. In contrast, when Advisors' preferences are misaligned, communication involves strategic incentives to exaggerate, and disclosure may affect trust. Pooling these environments may therefore mask important mechanisms, and the average payoffs in Figure 6 hide potential heterogeneity. We consequently analyze outcomes separately for aligned and misaligned Advisors to distinguish improvements driven by better coordination (in cases of aligned preferences) from those driven by reductions in strategic distortion (in cases of misaligned preferences).

Figure 7 presents the average payoffs for pairs with Advisors where there is perfect alignment of preferences between Advisors and their Clients (hence, we present the average payoffs at the pair level). We find that partner interaction slightly helps pairs when incentives are aligned and the disclosure of preferences is not possible (SN vs. PN, two-sided MW test, $p = 0.089$). Average payoffs of pairs with aligned Advisors are almost identical under mandatory (SM vs. PM, two-sided MW test, $p = 0.470$) and voluntary disclosure (SV vs. PV, two-sided MW test, $p = 0.434$).

Focusing on short-term interactions, we find that both types of disclosure significantly increase the payoffs of players when interests are aligned (SN vs. SM, two-sided MW test, $p = 0.007$; SN vs. SV, two-sided MW test, $p = 0.004$; SM vs. SV, two-sided MW test, $p = 0.581$). In partner interactions, this effect is weaker when comparing mandatory and no disclosure (PN vs. PM, two-sided MW test, $p = 0.078$), but preserved when comparing voluntary

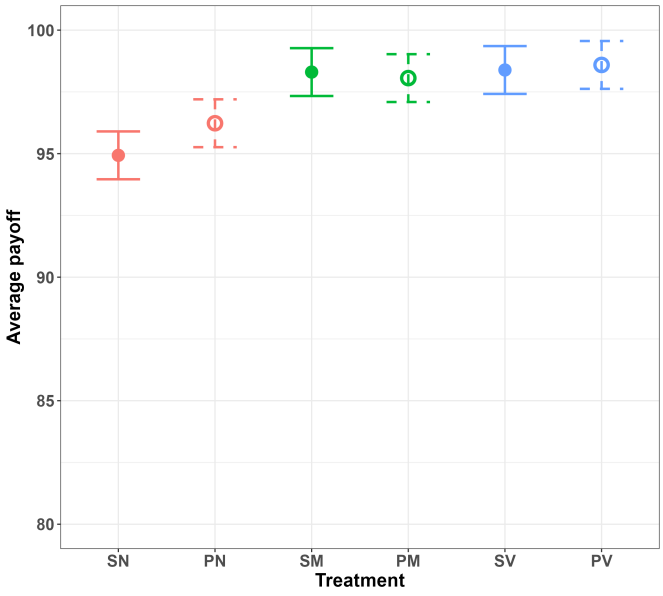


Figure 7: Average Payoffs of Aligned Advisors and Their Matched Clients Across Treatments. *Note:* The figure presents the average at the pair level as the outcomes for the two members are exactly the same when preferences are aligned.

with no disclosure (PN vs. PV, two-sided MW test, $p = 0.002$). The difference between PV and PM is not significant (two-sided MW test, $p = 0.644$).

Result 10 *When interests are aligned, relationship length does not matter irrespective of disclosure protocol. Voluntary disclosure helps aligned pairs, compared to no disclosure both in the short-term and partner setting. Mandatory disclosure helps these pairs only in short-term interactions.*

Figure 8 presents the average payoffs of Advisors and Clients when a conflict of interest was present (i.e., preferences were misaligned). We do not find any significant differences due to relationship length for either Advisors or Clients.

Turning to the effect of disclosure protocols within the short-term setting, we find that Advisors' average payoff is slightly lower when some form of disclosure is implemented, but the differences are small, and at most marginally significant (SN vs. SM, two-sided MW test, $p = 0.114$; SN vs. SV, two-sided MW test, $p = 0.052$; SM vs. SV, two-sided MW test, $p = 0.887$). Clients' average payoffs are relatively stable across disclosure treatments (SN vs. SM, two-sided MW test, $p = 0.514$; SN vs. SV, two-sided MW test, $p = 0.266$; SM vs. SV, two-sided MW test, $p = 0.443$).

In partner settings, misaligned Advisors are significantly hurt by mandatory disclosure compared to no disclosure (PN vs. PM, two-sided MW test, $p = 0.007$). Voluntary disclosure also has a negative effect on these Advisors but this effect is weaker (PN vs. PV, two-sided MW test, $p = 0.060$). There is no significant difference in payoffs for misaligned Advisors between the two types of disclosure protocols (PM vs. PV, two-sided MW test, $p = 0.799$). Clients in these interactions are not significantly benefiting from any form of disclosure (PN vs. PM, two-sided

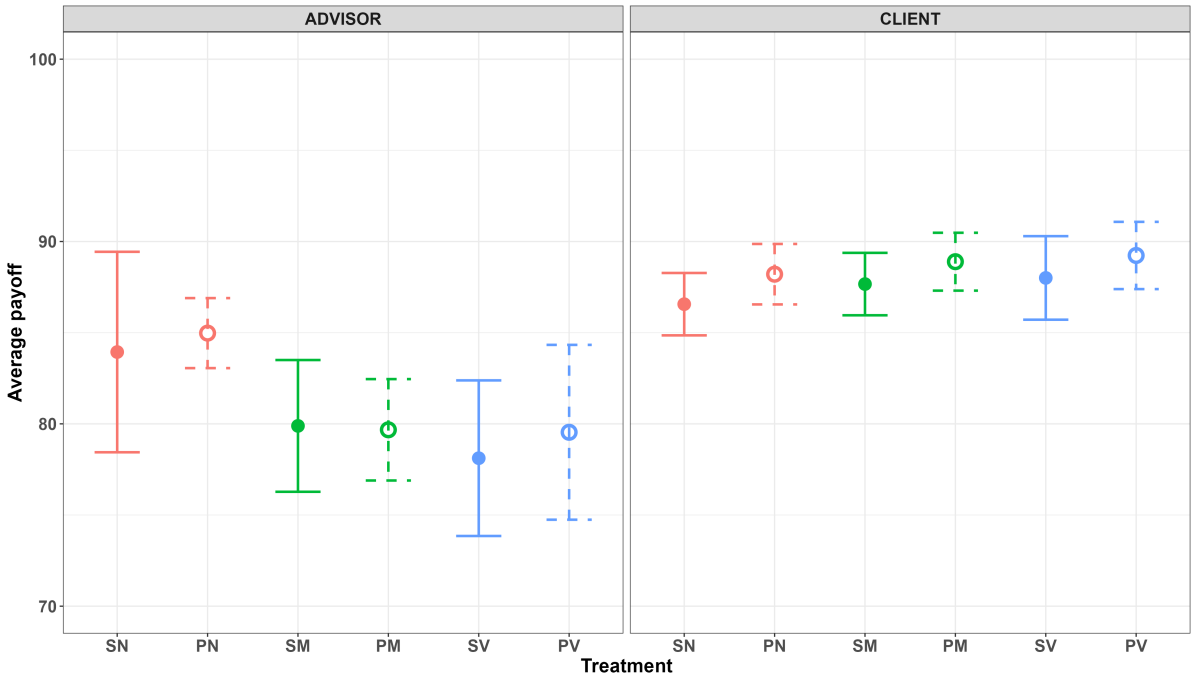


Figure 8: Average Payoffs of Misaligned Advisors and Their Matched Clients Across Treatments

MW test, $p = 0.514$; PN vs. PV, two-sided MW test, $p = 0.319$; PM vs. PV, two-sided MW test, $p = 0.671$).

Result 11 *When interests are conflicting, relationship length does not matter irrespective of disclosure protocol. Neither form of disclosure affects Clients' payoffs when their preferences are misaligned with those of the Advisors. However, these Advisors are worse off when disclosure is mandatory.*

Overall, the welfare results show that Clients benefit most consistently from voluntary disclosure, while mandatory disclosure yields benefits primarily when interests are aligned. When Advisors have misaligned preferences, disclosure does not improve Clients' outcomes and can reduce Advisors' payoffs. Repeated interaction generates welfare gains when no disclosure is possible, but these additional welfare gains are limited once disclosure is available. These patterns suggest that in our setting the increased trust does not necessarily results in higher welfare. This, however, might also be dependent on the chosen payoff function, as in our setting profit-sharing across agents is possible.

5 Conclusion and Discussion

In this paper we investigated how long-term interaction and different disclosure protocols affect behavior, and thereby welfare, in advisor–client relationships with a potential conflict of interest (COI). In our laboratory experiment, the Client's action determines the payoffs of both players, but only the Advisor knows the state of the world. Before the Client acts, the Advisor can send a costless cheap-talk message about the state. We vary the disclosure protocol for the Advisor's preferences (no, mandatory, and voluntary disclosure) and the length of the

interaction (short-term and partner matching). We find that Clients earn more under partner matching when they have no information about the Advisor's preferences. Moreover, both types of disclosure benefit Clients under short-term matching, whereas under partner matching only voluntary disclosure does. On average, Advisors are neither harmed nor helped by any of our treatment interventions. These welfare results stem from changes in the behavior of both Advisors and Clients. Advisors with a COI are less likely to exaggerate their messages in partner interactions, especially when they have voluntarily disclosed their COI. In response, Clients follow the advice more often after a voluntary disclosure of a COI in partner interactions, which also increases their decision accuracy. Furthermore, under voluntary disclosure, Clients interpret non-disclosure as a bad sign and "punish" the Advisor by discounting her message more often than when her preferences are exogenously kept hidden.

Our findings on mandatory disclosure contrast with theoretical (Li and Madarász, 2008) and experimental (Cain et al., 2005) work showing that mandatory disclosure can harm clients. However, our design differs from this literature in several respects. First, unlike most previous studies of mandatory disclosure, our design includes both aligned and misaligned Advisors with equal probability, rather than only one type, and Clients know this. As a result, an additional mechanism can play an important role in our setting: identifying the Advisor's preferences. We chose an environment with both types of Advisors because it better captures reality and, as an experimental feature, because it allows us to study voluntary disclosure decisions.

Second, Cain et al. (2005) and Cain et al. (2011) used an unbounded state space in which advisors did not know the true state either but had a better estimate of it, so clients did not know its upper bound. In those studies, the advisor's message may therefore also have served as an anchor for the client's action. In our design, participants know the state space, which allows Clients to disregard the message entirely and simply choose the expected value. While these differences from previous work are substantial, each design captures relevant real-world settings. An unbounded state space corresponds to situations such as bidding for a house or valuing other assets, whereas our bounded state space corresponds to decision problems with a finite set of options that can be ranked by their efficiency, such as choosing among investment options.

Finally, a few observations are noteworthy regarding interpreting the size of our welfare effects. Payoff differences across treatments are not straightforwardly interpretable, because their magnitude depends heavily on the payoff function we implemented. First, the constant of 100 in $\Pi^C = 100 - (s - a)^2$ was added only to avoid losses, so differences expressed relative to average earnings understate the effects. The relevant range is much narrower: a Client who ignores all messages and chooses $a \in \{5, 6\}$ earns 91.5 ECU in expectation, while a perfectly informed Client earns 100 ECU. Communication can therefore improve a Client's expected payoff by at most 8.5 ECU per period, and treatment differences should be judged against this range rather than against total earnings. Second, the quadratic loss function makes small errors cheap: an action that misses the state by one unit costs only 1 ECU, whereas a miss of four units costs 16 ECU. Behavioural changes that mainly affect small deviations, such as moderate exaggeration or Clients "rewarding" a misaligned Advisor by choosing slightly higher actions, therefore translate into small payoff differences, even when they substantially change trust and accuracy. This also explains why we observe sizeable effects on exaggeration, following behaviour and decision accuracy, but only modest and sometimes insignificant effects on payoffs. Third, with a different loss function, for instance a linear one or one that

penalises large errors more heavily, the same behaviour would map into different welfare differences. We therefore view the qualitative behavioural results as more robust than the exact magnitudes of the welfare effects.

Repeated interaction and reputation in advisor–client relationships with uncertain conflicts of interest remain understudied, and our results open up several avenues for future research. They also have potential implications for policy and practice. First, we show that long-term relationships have value beyond saving switching costs: our pairs built up trust and increased the Clients' decision accuracy.[15] Second, we demonstrate the value of voluntary disclosure and show that Clients perceive it as a signal of trustworthiness. Professionals in advisory roles, including doctors, may therefore want to consider sharing truthful information with their clients even when doing so does not intuitively seem to be in their own interest. Encouraging advisors to disclose, or informing them about the potential benefits of disclosure, might thus improve advisor–client relationships without the need for stricter regulatory interventions. We acknowledge that the laboratory differs from real-world settings in that our Clients knew that voluntary disclosure was possible, which is not necessarily the case in reality. Furthermore, our Advisors could always truthfully report the *absence* of a COI, which is harder to prove in reality than its *presence*. We leave experiments addressing these issues to future research.

[15] We refrained from modeling competition between Advisors and outside options for Clients, as this is beyond the scope of this project. Note, however, that our Clients could always disregard the message they received, which can be interpreted as taking an outside option.

References

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

Anagol, S., Cole, S., and Sarkar, S. (2017). Understanding the advice of commissions-motivated agents: Evidence from the indian life insurance market. *The Review of Economics and Statistics*, 99(1):1–15.

Balafoutas, L. and Kerschbamer, R. (2020). Credence goods in the literature: What the past fifteen years have taught us about fraud, incentives, and the role of institutions. *Journal of Behavioral and Experimental Finance*, 26:100285.

Barron, K. and Fries, T. (2024). Narrative persuasion. Technical report, WZB Discussion Paper.

Bayona, A., Peia, O., and Vlahu, R. (2026). Credit ratings and investments. *Management Science*.

Bénabou, R. and Laroque, G. (1992). Using privileged information to manipulate markets: Insiders, gurus, and credibility. *The Quarterly Journal of Economics*, 107(3):921–958.

Blume, A., Lai, E. K., and Lim, W. (2020). Strategic information transmission: A survey of experiments and theoretical foundations. In Capra, C. M., Croson, R. T. A., Rigdon, M. L., and Rosenblat, T. S., editors, *Handbook of Experimental Game Theory*, pages 311–347. Edward Elgar.

Bolton, P., Freixas, X., and Shapiro, J. (2012). The credit ratings game. *The Journal of Finance*, 67(1):85–112.

Cai, H. and Wang, J. T.-Y. (2006). Overcommunication in strategic information transmission games. *Games and Economic Behavior*, 56(1):7–36.

Cain, D. M., Loewenstein, G., and Moore, D. A. (2005). The dirt on coming clean: Perverse effects of disclosing conflicts of interest. *The Journal of Legal Studies*, 34(1):1–25.

Cain, D. M., Loewenstein, G., and Moore, D. A. (2011). When sunlight fails to disinfect: Understanding the perverse effects of disclosing conflicts of interest. *Journal of Consumer research*, 37(5):836–857.

Chen, P. and Richardson, M. (2018). Conflicted financial advice: disclosure revisited. *Applied Economics Letters*, 25(12):826–829.

Church, B. K. and Kuang, X. (2009). Conflicts of interest, disclosure, and (costly) sanctions: Experimental evidence. *The Journal of Legal Studies*, 38(2):505–532.

Crawford, V. P. and Sobel, J. (1982). Strategic Information Transmission. *Econometrica*, 50(6):1431–1451.

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

de Meza, D., Irlenbusch, B., and Reyniers, D. (2010). Disclosure, trust and persuasion in insurance markets. IZA Discussion Paper 5060, Institute of Labor Economics (IZA).

De Meza, D., Irlenbusch, B., and Reyniers, D. (2010). Disclosure, trust and persuasion in insurance markets. Technical report, IZA Discussion Papers.

Dranove, D. and Jin, G. Z. (2010). Quality disclosure and certification: Theory and practice. *Journal of Economic Literature*, 48(4):935–963.

Dulleck, U. and Kerschbamer, R. (2006). On doctors, mechanics, and computer specialists: The economics of credence goods. *Journal of Economic Literature*, 44(1):5–42.

Dulleck, U., Kerschbamer, R., and Sutter, M. (2011). The economics of credence goods: An experiment on the role of liability, verifiability, reputation, and competition. *American Economic Review*, 101(2):526–555.

Ely, J. C. and Välimäki, J. (2003). Bad reputation. *The Quarterly Journal of Economics*, 118(3):785–814.

Fischbacher, U. and Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, 11(3):525–547.

Gesche, T. (2021). De-biasing strategic communication. *Games and Economic Behavior*, 130:452–464.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, 95(1):384–394.

Gneezy, U., Kajackaite, A., and Sobel, J. (2018). Lying aversion and the size of the lie. *American Economic Review*, 108(2):419–453.

Grossman, S. J. (1981). The informational role of warranties and private disclosure about product quality. *The Journal of Law and Economics*, 24(3):461–483.

Hagenbach, J. and Perez-Richet, E. (2018). Communication with evidence in the lab. *Games and Economic Behavior*, 112:139–165.

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

Ismayilov, H. and Potters, J. (2013). Disclosing advisor's interests neither hurts nor helps. *Journal of Economic Behavior & Organization*, 93:314–320.

Jin, G. Z., Luca, M., and Martin, D. (2021). Is no news (perceived as) bad news? an experimental investigation of information disclosure. *American Economic Journal: Microeconomics*, 13(2):141–173.

Kawagoe, T. and Takizawa, H. (2009). Equilibrium refinement vs. level-k analysis: An experimental study of cheap-talk games with private information. *Games and Economic Behavior*, 66(1):238–255.

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

Kopányi-Peuker, A., Offerman, T., and Sloof, R. (2017). Fostering cooperation through the enhancement of own vulnerability. *Games and Economic Behavior*, 101:273–290.

Li, M. and Madarász, K. (2008). When mandatory disclosure hurts: Expert advice and conflicting interests. *Journal of Economic Theory*, 139(1):47–74.

Loewenstein, G., Cain, D. M., and Sah, S. (2011). The limits of transparency: Pitfalls and potential of disclosing conflicts of interest. *American Economic Review: Papers & Proceedings*, 101(3):423–428.

Loewenstein, G., Sah, S., and Cain, D. M. (2012). The unintended consequences of conflict of interest disclosure. *JAMA*, 307(7):669–670.

Mathis, J., McAndrews, J., and Rochet, J.-C. (2009). Rating the raters: Are reputation concerns powerful enough to discipline rating agencies? *Journal of Monetary Economics*, 56(5):657–674.

Milgrom, P. R. (1981). Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, 12(2):380–391.

Mimra, W., Rasch, A., and Waibel, C. (2016). Price competition and reputation in credence goods markets: Experimental evidence. *Games and Economic Behavior*, 100:337–352.

Monin, B. and Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology*, 81(1):33–43.

Morgan, J. and Stocken, P. C. (2003). An analysis of stock recommendations. *The RAND Journal of Economics*, 34(1):183–203.

Morris, S. (2001). Political correctness. *Journal of Political Economy*, 109(2):231–265.

Mullainathan, S., Noeth, M., and Schoar, A. (2012). The market for financial advice: An audit study. NBER Working Paper 17929, National Bureau of Economic Research.

Rabanal, J. P. and Rud, O. A. (2018). Does competition affect truth telling? an experiment with rating agencies. *Review of Finance*, 22(4):1581–1604.

Saccardo, S. and Serra-Garcia, M. (2023). Enabling or Limiting Cognitive Flexibility? Evidence of Demand for Moral Commitment. *American Economic Review*, 113(2):396–429.

Sah, S. (2019). Understanding the (perverse) effects of disclosing conflicts of interest: A direct replication study. *Journal of Economic Psychology*, 75:102118.

Sah, S. and Feiler, D. (2020). Conflict of interest disclosure with high-quality advice: The disclosure penalty and the altruistic signal. *Psychology, Public Policy, and Law*, 26(1):88–104.

Sah, S. and Loewenstein, G. (2014). Nothing to declare: Mandatory and voluntary disclosure leads advisors to avoid conflicts of interest. *Psychological Science*, 25(2):575–584.

Sah, S., Loewenstein, G., and Cain, D. (2019). Insinuation Anxiety: Concern That Advice Rejection Will Signal Distrust After Conflict of Interest Disclosures. *Personality and Social Psychology Bulletin*, 45(7):1099–1112.

Sah, S., Loewenstein, G., and Cain, D. M. (2013). The Burden of Disclosure: Increased Compliance with Distrusted Advice. *Journal of Personality and Social Psychology*, 104(2):289–304.

Sánchez-Pagés, S. and Vorsatz, M. (2007). An experimental study of truth-telling in a sender–receiver game. *Games and Economic Behavior*, 61(1):86–112.

Skreta, V. and Veldkamp, L. (2009). Ratings shopping and asset complexity: A theory of ratings inflation. *Journal of Monetary Economics*, 56(5):678–695.

Sobel, J. (1985). A theory of Credibility. *The Review of Economic Studies*, 52(4):557–573.

Stocken, P. C. (2000). Credibility of Voluntary Disclosure. *The RAND Journal of Economics*, 31(2):359–374.

Wang, J. T.-y., Spezio, M., and Camerer, C. F. (2010). Pinocchio's pupil: Using eyetracking and pupil dilation to understand truth telling and deception in sender-receiver games. *American Economic Review*, 100(3):984–1007.

Wilson, A. J. and Vespa, E. (2020). Information transmission under the shadow of the future: An experiment. *American Economic Journal: Microeconomics*, 12(4):75–98.

A Proofs of the theoretical predictions

Proof of Lemma 1 (Stage game under no disclosure): We start by showing that these strategies and beliefs constitute an equilibrium. Given the Client's strategy, the Advisor does not have an incentive to deviate from her strategy. In case the Advisor is unbiased, she achieves the highest possible welfare for $s \leq 6$. If $s > 6$, then the highest action that she can induce is still 6, even if she deviates to a message lower than 7. A biased Advisor also induces the highest possible action by choosing a message from $m \in \{7, 8, 9, 10\}$. Therefore neither type of advisor has an incentive to deviate from her strategy.

Given the Advisor's strategy, the Client achieves the highest payoff by following a message $m \leq 6$. The belief that these messages come surely from an unbiased Advisor is correct. For messages higher than 6, the Client cannot perfectly distinguish between the two types of Advisor, and they know that the message is uninformative about the higher states. The conditional probability that the Advisor is unbiased after observing a message $m \in \{7, 8, 9, 10\}$ is the following:

$$\Pr(\text{A unbiased}|m \in \{7, 8, 9, 10\}) = \frac{0.1 \cdot 0.5}{0.1 \cdot 0.5 + 0.25 \cdot 0.5} = \frac{2}{7}.$$

Given these beliefs, the expected state of the world is:

$$E(s|m \in \{7, 8, 9, 10\}) = \frac{2}{7} \cdot 8.5 + \frac{5}{7} \cdot 5.5 = \frac{89}{14} \approx 6.36$$

The best response to these beliefs from the Client is indeed to choose action 6. These show that these strategies and beliefs constitute an equilibrium.

As a next step we show that there is no semi-separating equilibrium that allows for more information transmission. Assume that there is an equilibrium where only the three highest states are pooled, and the Client reacts with $a = m$ for all $m < 8$. In that case $m \in \{8, 9, 10\}$ is uninformative about the three highest states, and gives the conditional probability of the advisor being unbiased as

$$\Pr(\text{A unbiased}|m \in \{8, 9, 10\}) = \frac{0.1 \cdot 0.5}{0.1 \cdot 0.5 + 1/3 \cdot 0.5} = \frac{3}{13}.$$

Given this probability, the expected state is $E(s|m \in \{8, 9, 10\}) = \frac{3}{13} \cdot 9 + \frac{10}{13} \cdot 5.5 = \frac{82}{13} \approx 6.31$ giving again a best reponse of $a = 6$. However, given this best response a biased Advisor has an incentive to deviate from her strategy, and choose $m = 7$ inducing a higher action $a = 7$. Therefore there exists no equilibrium in which more information is transmitted from the Advisor to the Client.

Q.E.D.

There are further equilibria of the stage game without disclosure that result in lower information transmission. An example would be where the Client has a threshold of $m = 5$ instead of $m = 6$, and follows only messages $m \leq 5$. For all other messages he chooses $a = 6$. The unbiased Advisor then randomly chooses from $m \in \{6, 7, 8, 9, 10\}$ when $s > 5$, and the biased Advisor randomly chooses from $m \in \{6, 7, 8, 9, 10\}$ for any s. In this case the Client's belief that the Advisor is unbiased after seeing a pooled message is $\Pr(\text{A unbiased}|m \in \{6, 7, 8, 9, 10\}) = 1/3$, and his expectation is $E(s|m \in \{6, 7, 8, 9, 10\} = 6.33$. Choosing $a = 6$ is indeed a best response then, and neither party has an incentive to deviate.

Another example is the babbling equilibrium where both types of Advisors choose a message randomly irrespective of the state, and the Client disregards this message and chooses $a = 5$ or $a = 6$. In this situation none of the parties have an incentive to deviate from their strategy.

Proof of Lemma 2 (Stage game under mandatory disclosure): When the Advisor in unbiased, then given her and the Client's strategy there is perfect information transmission, and both parties achieve the highest welfare. Therefore neither of them has an incentive to deviate.

When the Advisor is biased, then she does not have an incentive to deviate from her strategy, as the Client's action is irrespective of m. The Client does not have an incentive to deviate either, because the message is uninformative resulting in an expected state of $E(s) = 5.5$. $a = 5$ or $a = 6$ are best responses against it.

It remains to show that there is no other equilibrium in which the biased Advisor transmits some information. Assume that there is a state s' in which the Advisor can transmit some information credibly, and the Client reacts to that by choosing his action accordingly. This is only profitable for the Advisor, if she can earn higher payoffs than in the babbling equilibrium. However, if there is a state where this is possible, then the Advisor has an incentive to always report that state irrespective of the actually realized state. If that is the case, then the message becomes again uninformative leading to $a = 5$ or $a = 6$. Therefore there is no equilibrium in which the biased Advisor transmits some information.

Q.E.D.

Similarly to the equilibrium under No disclosure, we can also construct equilibria with lower information transmission. An example is the babbling equilibrium where both types of Advisor send a random message for all states, and the Client disregards completely all messages he receives. Neither the Advisor nor the Client can improve their situation by unilaterally deviating from these strategies.

B Additional Analyses

Table 3: Effect of Disclosure on Truthful Messaging by Unbiased Advisors

	Forced (Non-)Disclosure (Forced)	Voluntary Disclosure (SV & PV)
Disclosed	0.051**	0.129***
	(0.025)	(0.043)
Partner interaction	-0.004	0.021
	(0.029)	(0.018)
Constant	0.878***	0.833***
	(0.018)	(0.050)
Observations	2740	1319

Notes: Linear probability models. The dependent variable equals one if the advisor sent a truthful message. The sample is restricted to unbiased advisors. The first column compares mandatory disclosure to no disclosure (SN/PN vs. SM/PM). The second column uses only voluntary disclosure treatments and compares periods in which the advisor disclosed their type to those in which they did not. Robust standard errors clustered at the matching-group level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 4: Relationship Length and Exaggeration by Biased Advisors

	Dependent variable: Exaggeration		
	All Treatments	Voluntary: SV & PV	
		(Disclosed)	(Not Disclosed)
Partner interaction	-0.058**	-0.182***	0.101
	(0.025)	(0.045)	(0.071)
Mandatory disclosure	-0.035		
	(0.027)		
Voluntary disclosure	-0.125**		
	(0.058)		
Constant	0.744***	0.612***	0.664***
	(0.019)	(0.052)	(0.079)
Clustered SE	Yes	Yes	Yes
Observations	4041	852	457

Notes: Linear probability models. The dependent variable equals one if the advisor exaggerated the message. The sample includes only biased advisors. Column (1) includes all treatments. Mandatory disclosure corresponds to treatments with required disclosure, while voluntary disclosure indicates treatments where advisors could choose whether to disclose. Column (2) restricts the sample to voluntary disclosure treatments and periods in which the advisor disclosed their bias. Partner interaction indicates repeated interaction. Robust standard errors clustered at the matching-group level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 5: Partner Interaction and Voluntary Disclosure Effects on Following Behavior (Undisclosed Bias)

	Dependent variable: Follow		
	(1) Partner	(2) Voluntary	(3) Partner × Voluntary
Partner interaction	0.097*		0.109
	(0.057)		(0.073)
Voluntary disclosure		-0.153***	-0.145*
		(0.045)	(0.079)
Partner × Voluntary			-0.027
			(0.096)
Constant	0.484***	0.561***	0.507***
	(0.048)	(0.032)	(0.054)
Observations	3331	3331	3331

Notes: Linear probability models. The dependent variable equals one if the Client followed the advisor's message. The sample includes observations in which bias was not disclosed. Partner interaction indicates repeated interaction. Voluntary disclosure setting equals one in voluntary disclosure treatments. Standard errors clustered at the matching-group level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 6: Partner Interaction and Client Responses when Bias is Undisclosed

	Discount	Reward
Panel A: No Disclosure Setting		
Partner interaction	-0.054	-0.055**
	(0.091)	(0.023)
Constant	0.402***	0.091***
	(0.063)	(0.019)
Panel B: Voluntary Disclosure Setting		
Partner interaction	-0.146**	0.064*
	(0.058)	(0.037)
Constant	0.583***	0.056***
	(0.062)	(0.015)
Observations	2736	595

Notes: Linear probability models. The dependent variables equal one if the Client discounted or rewarded the Advisor's recommendation. The sample includes observations in which bias was not disclosed. Panel A restricts the sample to treatments without voluntary disclosure. Panel B restricts the sample to voluntary disclosure treatments. Standard errors clustered at the matching-group level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 7: Partner Interaction and Client Responses when Bias is Disclosed (Biased Advisors)

	Follow	Discount	Reward
Panel A: Mandatory Disclosure			
Partner interaction	0.063	-0.107	0.044
	(0.060)	(0.083)	(0.034)
Constant	0.233***	0.697***	0.070***
	(0.033)	(0.039)	(0.011)
Observations	1366	1366	1366
Panel B: Voluntary Disclosure			
Partner interaction	0.142***	-0.219***	0.078**
	(0.052)	(0.068)	(0.035)
Constant	0.280***	0.631***	0.089***
	(0.037)	(0.056)	(0.032)
Observations	852	852	852

Notes: Linear probability models. The dependent variables equal one if the Client followed, discounted, or rewarded the Advisor's recommendation. The sample includes only observations with biased advisors and in which the bias was disclosed. Panel A restricts the sample to mandatory disclosure treatments. Panel B restricts the sample to voluntary disclosure treatments. Partner interaction indicates repeated interaction. Standard errors clustered at the matching-group level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 8: Effect of Advisor Bias and Disclosure on Decision Accuracy

	Dependent variable: Accuracy	
	(1)	(2)
Bias drawn	-0.601***	-0.590***
	(0.018)	(0.018)
Bias disclosed		0.133***
		(0.018)
Constant	0.779***	0.696***
	(0.023)	(0.023)
Observations	8100	8100

Notes: Linear probability models. The dependent variable equals one if the client's decision matches the true state. Bias drawn indicates that the advisor had a conflict of interest. Bias disclosed indicates that the conflict of interest was revealed to the client. Standard errors clustered at the matching-group level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 9: Effect of Partner Interaction on Decision Accuracy

	Dependent variable: Accuracy	
	(1) Overall	(2) Voluntary Disclosure
		Bias Discl. & Biased
Partner interaction	0.041*	0.121**
	(0.023)	(0.049)
Constant	0.459***	0.181***
	(0.021)	(0.022)
Sample	Full sample	Voluntary (bias = 1, disclosed = 1)
Observations	8100	852

Notes: Linear probability models. The dependent variable equals one if the client's decision matches the true state. Partner interaction indicates repeated interaction. Column (2) restricts the sample to voluntary disclosure treatments in which the advisor was biased and disclosed the conflict of interest. Standard errors clustered at the matching-group level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 10: Voluntary Disclosure Environment and Decision Accuracy

	Dependent variable: Accuracy	
	(1) Bias Disclosed	(2) Bias Not Disclosed
Voluntary disclosure	0.126***	-0.158***
	(0.024)	(0.033)
Partner interaction	0.019	0.084**
	(0.029)	(0.036)
Constant	0.491***	0.358***
	(0.029)	(0.025)
Sample	Bias disclosed	Bias not disclosed
Observations	4769	3331

Notes: Linear probability models. The dependent variable equals one if the client's decision matches the true state. Voluntary disclosure environment equals one in voluntary disclosure treatments. Partner interaction indicates repeated interaction. Column (1) includes only observations where bias was disclosed; column (2) includes only observations where bias was not disclosed. Standard errors clustered at the matching-group level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 11: Disclosure Protocols and Decision Accuracy

	Dependent variable: Accuracy		
	(1) Voluntary vs None	(2) Mandatory vs None	(3) Voluntary vs Mandatory
Voluntary disclosure	0.141***		0.041
	(0.028)		(0.026)
No disclosure		-0.101***	
		(0.020)	
Constant	0.400***	0.500***	0.500***
	(0.023)	(0.030)	(0.030)
Comparison	V vs N	M vs N	V vs M
Observations	5364	5472	5364

Notes: Linear probability models. The dependent variable equals one if the client's decision matches the true state. Column (1) compares voluntary disclosure and no-disclosure environments (SV, PV vs. SN, PN). Column (2) compares mandatory disclosure and no-disclosure environments (SM, PM vs. SN, PN). Column (3) compares voluntary and mandatory disclosure environments (SV, PV vs. SM, PM). Standard errors clustered at the matching-group level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

C Experimental instructions and screenshots

C.1 Instructions Partner Voluntary treatment

This appendix contains the experimental instructions for the PV treatment. The other instructions are similar, and can be found in the OSF-repository together with the preregistration.

PAGE 1

Welcome and thank you for participating in this experiment on decision-making. Please read the following detailed instructions carefully. On your desk you will find a short summary of these instructions that you may refer to at any time during the experiment.

Your earnings will depend partly on your decisions, partly on the decisions of others, and partly on chance. Pay attention and follow the instructions closely, as we will explain how you will earn money and how your earnings will depend on the choices that you make. Irrespective of this, you will receive an additional **10 EURO for showing up on time and completing the experiment**. All payments will take place via online transfers.

All your decisions are anonymous, so your identity will be kept secret at all times.

Throughout the whole experiment you must remain seated and mobile phones and other electronic devices must be switched off. If there are any questions, please raise your hand and an experimenter will come to answer your questions in private.

This experiment consists of two parts. Next, we will explain to you the first part of the experiment. You will receive instructions for the second part after the first part ends. The currency in the experiment is ECU (experimental currency unit) which will be exchanged to EURO using the following exchange rate: **1 ECU = 0.05 EURO**.

PAGE 2

Part 1 - General description

The first part of the experiment consists of several periods where participants interact in pairs. Each pair consists of an Advisor and a Client and you will play either in the role of the Advisor or that of the Client. Which role you are playing is randomly determined, and you will learn it after everyone goes through these instructions and before the first period begins. There is an equal probability that you will end up playing in either role. Your role stays the same for the entire experiment.

Throughout this part, the Client needs to make a choice that will influence the payoff of both the Client and the Advisor. This choice involves choosing an integer number from 1 to 10. The Client's payoff is the highest if their choice matches a randomly drawn integer number (also from 1 to 10). However, when making this choice, the Client does not know this randomly drawn number. The Advisor, instead, knows this extra piece of information, and can send a message to the Client. After observing the message, the Client will make a choice that will affect both players' payoff. In what follows we describe both roles in detail.

Information and task of the Advisor

At the beginning of each period the Advisor observes on their screen, an integer number, randomly drawn from 1 to 10. We will refer to this as the "**random number**" from now on. Then, the Advisor decides what **message** to send to the Client. The message can be any integer number from 1 to 10. The Advisor's message does not affect either their or the Client's payoff. The Advisor's payoff depends on the number chosen by the Client. We will refer to this as the "**chosen number**" from now on.

In each period the Advisor can be one of two types: type A or type B. Being either type is equally likely; that is, each type occurs with 50% probability, and this is determined anew in every period. This means that **the Advisor's type (A or B) may change from period to period**.

A **type A Advisor** has the **same preferences as the Client**: they want the Client to choose a number equal to the realized random number, in order to achieve the highest possible payoff. In each period, a type A Advisor's payoff (in ECU) is computed by the following function:

$$\text{Advisor A's Payoff} = 100 - (\text{random number} - \text{chosen number})^2$$

A **type B Advisor**, on the other hand, earns the **highest payoff if the Client chooses the maximum number, i.e. 10**, irrespective of what the random number is. In each period, a type B Advisor's payoff (in ECU) is computed by the following function:

$$\text{Advisor B's Payoff} = 100 - (10 - \text{chosen number})^2$$

When observing the random number, the Advisor also observes their type, and therefore the function that determines their payoff in that period. Thus, **the Advisor knows exactly how their payoff depends on the random and the chosen numbers**.

When sending the message, the Advisor also has the **possibility to inform the Client about whether they are of type A or type B**. The information the Advisor can send to the Client about their type is dependent on their type: it they are of type A, they can only reveal that they are of type A; if they are of type B, they can only reveal that they are of type B.

If the Advisor decides to inform the Client about their type, the Client observes the Advisor's type together with the Advisor's message, before choosing a number. If the Advisor decides not to inform the Client about their type, the Client observes only the Advisor's message before choosing a number.

Information and task of the Client

The Client's only task is to choose an integer number from 1 to 10 after observing the message sent by the Advisor. **Only if the Advisor decides to inform the Client about their type, the client will also know, with certainty, whether the Advisor is of type A or type B when choosing a number.** Otherwise, the Client will only know that both types can occur with the same probability in every period. The Client's payoff in each period depends both on the chosen number and the random number in the following way (in ECU):

$$\text{Client's Payoff} = 100 - (\text{random number} - \text{chosen number})^2$$

Thus, the Client's payoff is always the same as the payoff of type A Advisor. **Feedback at the end of each period**

After each period, if you are the Advisor, you will learn the Client's chosen number and your own payoff for that period.

If you are the Client, you will learn your own payoff for that period.

PAGE 3

Length of the experiment and matching of Advisors and Clients

The game described above is repeated multiple times (periods) in different blocks. How long the experiment will take is unknown to you and depends on a random draw in each period (different than the random number). In particular, in each period we draw a random number between 1 and 100. If this number is less than or equal to 90, you will play another period. If the number is larger than 90, the block will stop, and a new block will start (provided there is enough time left in the experiment). However, in the first 10 periods of each block you will not be informed about the outcome of the random draws. After period 10 we will inform you about all the random draws for the previous periods. If any of those draws are higher than 90, then the block is terminated, and only the periods up to and including the one with the random draw higher than 90 will potentially count towards your final earnings (as described in the next section of these instructions). If all the draws are at most 90, you will play further periods, but then you will immediately learn the result of the random draw after each period ends.

In each period of each block, you are **matched with the same participant.** Only when a new block starts, you will be matched with a different participant. You are never matched with the same participant across two

consecutive blocks. Also, you never receive any personal identification of the other participants.

Recall also that in each period the type of the Advisor is determined again with a random draw independent from the previous one. Furthermore, the random number the Advisor observed at the beginning of each period is determined for each period, independently from the one in the previous period. This means that the type of the Advisor or the random number in one period does not reveal any information about the Advisor type or the random number in the next period.

Payment from Part 1

After the end of Part 2 of the experiment, you will get feedback on your final earnings from this experiment. To determine your Part 1 earnings, the computer will randomly choose three periods from all the periods eligible for payment in Part 1 (see information above), add up the ECU earned in these periods and apply the 1 ECU = 0.05 EURO exchange rate. This will be added to your Part 2 earnings and the show-up fee to determine your final earnings.

C.2 Screenshots Partner Voluntary treatment

This subsection contains examples of screenshots in the Partner Voluntary treatment.

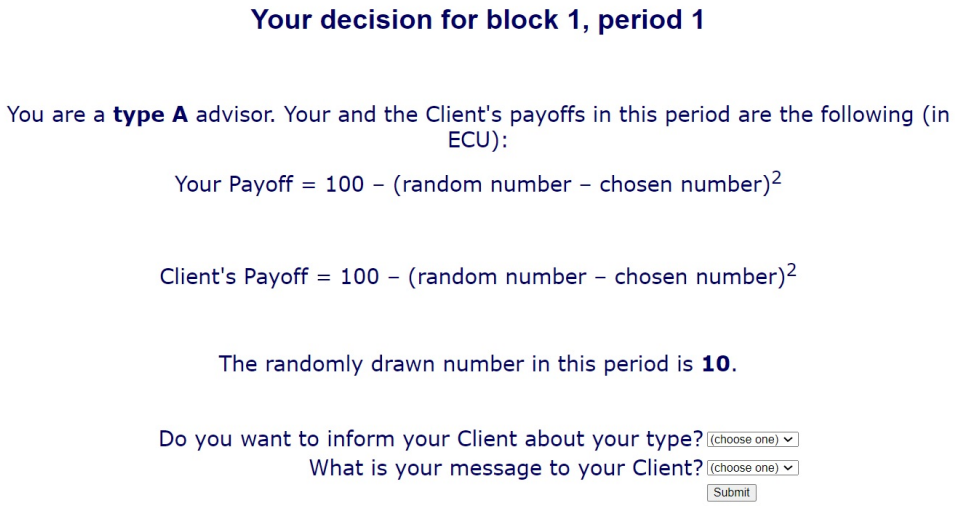


Figure C.1: Advisor's decision

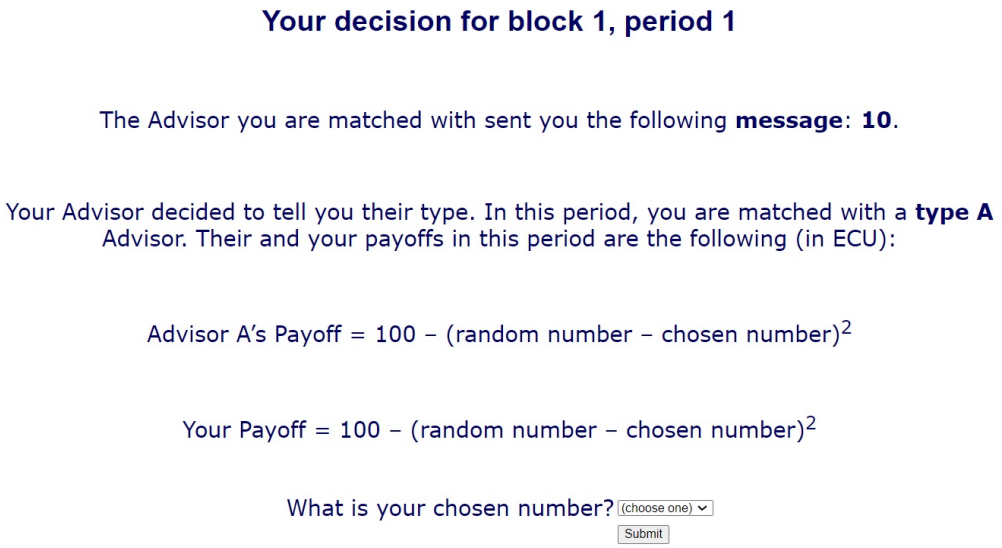


Figure C.2: Client's decision when Advisor disclosed her type

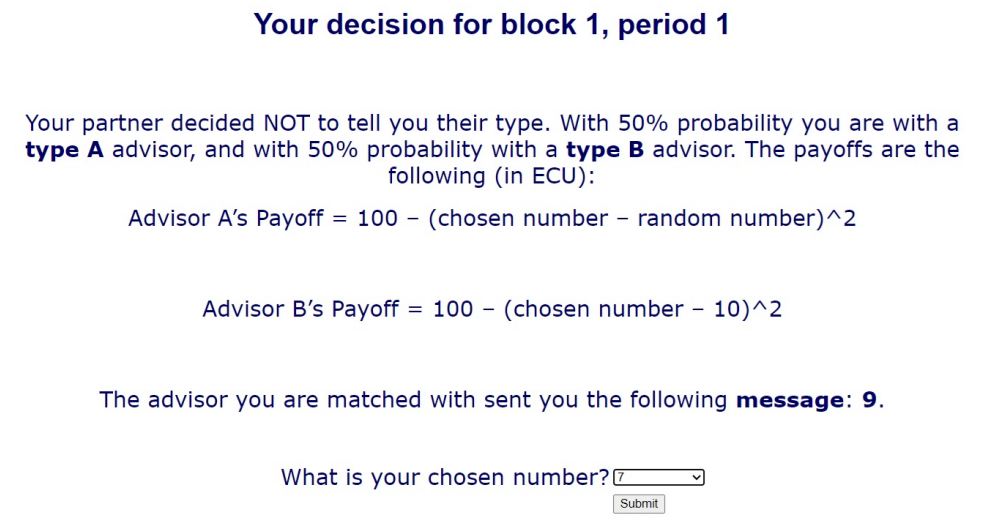


Figure C.3: Client's decision when Advisor did not disclose her type

C.3 Belief elicitation and lying preferences

In this section we present the instructions to the second part of the experiment. The different roles received different belief elicitation questions which we report for transparency. Participants have not seen any treatment or role labels.

PART 2 - PAGE 1

In this part, we are first interested in your beliefs corresponding to different events that might have occurred

during Part 1 of the experiment. There are no right or wrong answers so please answer as truthfully as you can. You will receive 1 EURO for completing this belief reporting task.
Recall the payoff function of each Advisor type:

$$\text{Advisor A's Payoff} = 100 - (\text{random number} - \text{chosen number})^2$$

$$\text{Advisor B's Payoff} = 100 - (10 - \text{chosen number})^2$$

[Advisors - No disclosure]
What percentage of Clients in this experiment do you think chose a number equal to the Advisor's message (i.e. % of Clients following the Advisor's message)?
Please fill in a percentage from 0 to 100 for each potential message of the Advisor. *[potential messages: 2,4,6,8,10]*

[Advisors - Mandatory disclosure]
Consider the cases were the Advisor was of type B.
What percentage of Clients in this experiment do you think chose a number equal to a type B Advisor's message (i.e. % of Clients following the type B Advisor's message)?
Please fill in a percentage from 0 to 100 for each potential message of the Advisor: *[potential messages: 2,4,6,8,10]*

[Advisors - Voluntary disclosure]
[The two questions are displayed in a random order randomized on the individual level.]
Consider the cases were the Advisor did not inform their Client about their type (i.e., Advisor's type is unknown to the Client).
What percentage of Clients in this experiment do you think chose a number equal to the Advisor's message (i.e. % of Clients following the message of an Advisor of unknown type)?
Please fill in a percentage from 0 to 100 for each potential message of the Advisor. *[potential messages: 2,4,6,8,10]*

Consider the cases were the Advisor informed their Client about their type and they were of type B (i.e., the Client knows that the Advisor is of type B).
What percentage of Clients in this experiment do you think chose a number equal to a type B Advisor's message (i.e. % of Clients following the type B Advisor's message)?
Please fill in a percentage from 0 to 100 for each potential message of the Advisor. *[potential messages: 2,4,6,8,10]*

[Clients - No disclosure]
What percentage of Advisors' messages in this experiment do you think were equal to the random number?
Please fill in a percentage from 0 to 100 for each potential random number the Advisor might have observed. *[potential random numbers: 2,4,6,8,10]*

[Clients - Mandatory disclosure]

Consider the cases were the Advisor was of type B.
What percentage (%) of Advisors' messages in this experiment do you think were equal to the random number?
Please fill in below a percentage from 0 to 100 for each of the listed potential random numbers the Advisor might have observed. *[potential random numbers: 2,4,6,8,10]*

[Clients - Voluntary disclosure]
[The two types of questions are displayed in a random order randomized on the individual level. The second type of question consisting of two questions (about type A and type B advisors) are always displayed in the same order.]
Consider the cases were the Advisor informed their Client about their type and they were of type B (i.e., the Client knows that the Advisor is of type B).
What percentage (%) of Advisors' messages in this experiment do you think were equal to the random number?
Please fill in below a percentage from 0 to 100 for each of the listed potential random numbers the Advisor might have observed. *[potential random numbers: 2,4,6,8,10]*

Consider the cases in this experiment where the Advisor observed they were of type A. In what percentage (%) of these cases do you think the Advisor decided to inform their Client about their type?
Consider the cases in this experiment where the Advisor observed they were of type B. In what percentage (%) of these cases do you think the Advisor decided to inform their Client about their type?

PAGE 2

In this task the computer will randomly assign you a number, 1, 2, 3, 4, 5, or 6. Each number has the same probability to be selected. You will then have to claim what number was assigned by the computer. You may claim that the assigned number was either "1", "2", "3", "4", "5" or "6", regardless of the actual number assigned.

You will first determine in advance what number you intend to claim for each of the 6 numbers the computer might assign. Later, the computer will randomly select an actual assigned number and only your claim corresponding to that number will be used to determine your earnings from this task.

Your earnings
You will receive 100 points, plus 20 times the number claimed for the actual assigned number. Note that for this task, 10 points = 0.05 EURO.

Thus, regardless of the actual assigned number, you will receive 120 points if you claim that the assigned number is "1"; 140 points if you claim that the assigned number is "2"; 160 points if you claim that the assigned number is "3"; and so on up to 220 points if you claim that the assigned number is "6".

When you are ready to start, please proceed to the next page.

QUESTIONNAIRE

Generally speaking, would you say that most people can be trusted or that you cannot be too careful in dealing with people?

- Most people can be trusted
- You cannot be too careful in dealing with people

In a sentence or two, could you explain how you made your decisions in Part 1 of this study? *[open question]*
Demographics: gender, age, field of study *[each with multiple choice, see preregistration]*