Journal of Banking and Finance 148 (2023) 106750

Contents lists available at ScienceDirect

Journal of Banking and Finance

journal homepage: www.elsevier.com/locate/jbf

The Banker's oath and financial advice

Utz Weitzel[a,b], Michael Kirchler[c,*]

[a] Vrije Universiteit Amsterdam & Tinbergen Institute, De Boelelaan 1105, Amsterdam, 1081HV, Netherlands
[b] Radboud University, Institute for Management Research, Heyendaalseweg 141, Nijmegen 6525AJ, Netherlands
[c] University of Innsbruck, Department of Banking & and Finance, Universitätsstrasse 15, Innsbruck 6020, Austria

ARTICLE INFO

Article history:
Received 8 September 2021
Accepted 15 December 2022
Available online 21 December 2022

JEL classification:
C92
D84
G02
G14

Keywords:
Experimental finance
Audit study
Banker's oath
Nudges
Financial advice

ABSTRACT

Financial misbehavior is widespread and costly. The Dutch government legally requires every employee in the financial sector to take a Hippocratic oath, the so-called "banker's oath." We investigate whether nudges that (in)directly remind financial advisers of their oath affect their service. In a large-scale audit study, professional auditors confronted 201 Dutch financial advisers with a conflict of interest. We find that when auditors apply a nudge that directly refers to the banker's oath, advisers are less likely to prioritize bank's interests. In additional prediction tasks, we find that Dutch regulators expect stronger effects of the oath than observed.

1. Introduction

Misbehavior in the financial sector is widespread and costly (Reurink, 2018). High-profile scandals, such as the LIBOR manipulation, the Wells Fargo account fraud, and the recent, global money-laundering scandal, make the news, but these are the exceptional cases of a much broader phenomenon that plays out at many levels and functions in the financial industry, including financial advice. In the period from 2005 to 2015, seven percent of financial advisers in the US are found to have misconduct records, reaching more than 15 percent at some of the largest firms (Egan et al., 2019).

The market for financial advice is particularly susceptible to misbehavior. First, financial advice resembles a credence good (Darby and Karni, 1973; Dulleck and Kerschbamer, 2006): bank customers often cannot fully appreciate the quality of the offered financial products or services, neither ex ante nor ex post. Second, financial advisers are typically subject to a conflict of interest: on the one hand they are supposed to provide advice in the customers' best interest, but on the other hand financial advisers are also expected to increase the profits of their employer. With regard to the latter, advisers may face commissions, fees, or key performance indicators that can create incentives for misbehavior (Inderst and Ottaviani, 2009; Inderst, 2010; Inderst and Ottaviani, 2012b), such as overtreatment and overcharging. It is therefore not surprising that this conflict of interest is often solved in the favor of the employer (Mullainathan et al., 2012; Fecht et al., 2018; Hoechle et al., 2018).

There is no easy remedy or golden bullet. Many policy interventions are debated and crucially depend on specific product features and on particular market channels (Inderst and Ottaviani, 2012a). Moreover, they do not necessarily increase customers' welfare in equilibrium (Chang and Szydlowski, 2020). The Netherlands therefore introduced, next to other policy interventions, a rather unusual and novel instrument: the so-called "banker's oath". Modelled after the Hippocratic oath for medical doctors, the Netherlands was worldwide the first country to impose by law an oath of ethics in the financial sector. Since January 1, 2015, every employee working in the financial sector in the Netherlands is legally required to take the following oath (and sign it) in a special ceremony arranged by the employers:

"I swear / promise that, within the boundaries of my function in the banking sector, I will:

- execute my function ethically and with care;
- draw a careful balance between the interests of all parties associated with the business, being the customers, shareholders, employees and the society in which the business operates;

* Corresponding author.
E-mail address: michael.kirchler@uibk.ac.at (M. Kirchler).

https://doi.org/10.1016/j.jbankfin.2022.106750

- when drawing that balance, make the customer's interests central;
- will comply with the laws, regulations and codes that apply to me;
- will keep confidential that which has been entrusted to me;
- will not abuse my knowledge;
- will act openly and accountably, knowing my responsibility to society;
- will make every effort to improve and retain trust in the financial sector.

So help me God! / This I pledge and promise!"

With this pledge the employee commits to put the customer's interest first (Loonen and Rutgers, 2017) and to comply with certain rules of conduct. Employees are personally responsible for compliance and can be held accountable for non-compliance.[1]

Five years after the general introduction of banker's oath, this study investigates whether nudges that directly or indirectly remind bank employees of their oath affect the quality of financial advice in a conflict of interest. To measure the (mis)behavior of financial advisers, we implement a large-scale audit study, where 51 professional auditors, disguised as normal customers, visited 201 bank branches in the Netherlands and confronted advisers with a pre-tested scenario. Audit studies are particularly suited to elicit socially undesirable behavior, as advisers are unaware that they are being studied. In economics, audit studies generated ground-breaking insights that would have been difficult to gather with other approaches. For instance, Bertrand and Mullainathan (2004) and Carlsson and Rooth (2007) show evidence of racial and ethnic discrimination in the labor market by sending out resumes to employers. In finance, audit studies have been used to analyze several topics, such as compliance with the prohibitions of setting up anonymous shell companies without proof of identity (Sharman, 2010), unethical portfolio advice to retail investors (Mullainathan et al., 2012), and the effects of ethics training in a bank (Harms, 2018).

In our study, financial advisers were randomly assigned to three different treatments. In the *control treatment*, CONTROL, the auditors were trained to impersonate a scenario where they wanted to take out a car loan for € 8,000 but also had savings of € 12,000 (without specific plans what to do with that money). Given that loans generate fees for banks and that interest rates for consumer loans are significantly higher than for savings, the scenario constituted a conflict of interest between protecting the customer from additional costs like fees and interest payments and selling a product for the benefit of the bank.[2] Harms (2018) used a similar scenario in an audit study to test the impact of a financial advisers' ethics training program on their advice-giving behavior. Auditors were blinded to the purpose of our study and did not have an account at the visited bank so that advisers were unable to look up more information. Directly after the visit, the auditors filled out a standardized questionnaire and recorded, in addition to basic demographic information, how strongly the adviser recommended to take out a loan and how strongly they recommended to use the savings.

In a second treatment, referred to as *direct nudge* (DIRECT), auditors administered a nudge that directly reminded financial advisers of the oath they took: at the beginning of the personal consultation, auditors mentioned that they have heard of a banker's oath and explicitly asked the adviser about the purpose of the oath. After the answer of the adviser, auditors proceeded with the scenario explained above in the control treatment.

In a third treatment, INDIRECT, auditors implemented an *indirect nudge* that reminded advisers of the most central element of the banker's oath, customer's interests (Loonen and Rutgers, 2017), without explicitly referring to the banker's oath. Specifically, auditors mentioned that they come from another bank which, according to their opinion, cares more about their own profits than about their clients. Then they asked advisers about how their bank protects customers' interests. After the answer, auditors proceeded as in CONTROL and DIRECT. With this treatment we attempt to disentangle effects that are linked to the pledge's central fiduciary element, putting customer's interests first, from the entirety of the oath, with all its other rules of conduct, underlying norms and values, and ceremonious character. This treatment extends to other customer-centered, professional codes of conduct that employers of many banks are asked to sign in acknowledgment, but without the solemnity of an oath (see Boatright (2013) for a detailed discussion).

Field experiments run the risk to produce Null results and are exposed to hindsight bias (DellaVigna et al., 2019). Therefore, before the results of the audit study were known, we sent out two online surveys with an incentivized prediction task that elicited the expected outcome of the audit study. One survey was sent to Dutch experts working in financial regulation and policy making (EXPERTS; N=122), and the same prediction task was administered to a representative sample of the Dutch working population (CUSTOMERS; N=502) as an approximation for bank customers who potentially apply for a car loan (for brevity, henceforth referred to as 'customers', 'potential customers' or the sample CUSTOMERS.) The priors of these two groups provide us with ex ante treatment averages for power tests and for an ex post comparison with the actual results without hindsight bias ("I knew that already!").

The results of the audit study show that, without any intervention, nearly half of all financial advisers (46.3%) prioritize loans in their recommendations. We find that direct nudges, reminding financial advisers of their oath, significantly decrease the likelihood for prioritizing loans by 16.4 percentage points (to 29.9% in Treatment DIRECT). We therefore detect a clear and substantial treatment effect of the direct nudge on the prioritization of product sales. Interestingly, this treatment effect is only caused by direct nudges, but not by indirect nudges (Treatment INDIRECT), suggesting that the mentioning of the oath triggers more than just an increase in the salience of customer interests. When we analyze the net strength of financial advice, we find that the direct nudge (and to some extent also the indirect nudge) increases neutral advice – where advisors neither favor loans nor savings – thereby reducing the frequency of recommendations that prioritize product sales.

When focusing on the expectations of the experts, we find that they predict significant treatment effects for both nudges, direct and indirect. Without intervention, experts expect financial advisors to prioritize loans much more than they actually do in the field (63.1% versus 46.3%, respectively). Both the experts and the customers correctly predict that direct nudges reduce financial advice that prioritizes product sales. However, experts (not customers) are wrong in expecting the same effect from indirect nudges. Overall, the predictions by the experts are less accurate (and more optimistic with regard to the effectiveness of direct nudges) than those by the customers. Power calculations show that the sample size of the audit study is large enough to comfortably rule out false negatives (type II errors) of the treatment effect sizes that were predicted by the experts.

[1] In the US, investment advisers are subject to the fiduciary standard, which requires them to place their clients' interests ahead of their own. This requirement is similar to the central element of the banker's oath. For more details on the banker's oath and its implementation by the independent Foundation for Banking Ethics Enforcement, see https://www.tuchtrechtbanken.nl/en.

[2] Since 2013, financial advisers and intermediaries in the Netherlands may not be compensated through commissions. Although, there are no direct monetary (commission-based) incentives for advisers to recommend a loan, there exist indirect benefits, such as satisfying key performance indicators, and career progress.

To the best of our knowledge, this is the first study that applies a randomized controlled trial (RCT) to estimate the causal effect of nudges and Hippocratic oaths on financial advice. In doing so we build on a nascent literature of experimental evidence from the field on the determinants of ethical financial advice. In an early audit study, Oehler and Kohlert (2009) document that auditors who impersonate greater financial sophistication receive better advice. Mullainathan et al. (2012) show that investment advisers fail to de-bias their clients and often even reinforce biases in order to advance advisers' personal interests. Anagol et al. (2017) conducted a series of audit studies to evaluate the quality of life insurance advice. They find that advisers overwhelmingly recommend unsuitable, strictly dominated products with high fees for the agent. In a related study, Harms (2018) implemented an ethics training program in a Dutch bank and subsequently analyzed its effect on financial advice in the field with an audit study. The authors report a substantial amount of unethical advice, but do not find any treatment effects of the program itself. We add to this literature by outlining that nudges addressing the banker's oath can help to decrease the likelihood that recommendations prioritize product sales (i.e., loans). In addition, we contribute by showing that experts' beliefs about loan-provision were exaggerated, given the audit data.

Our results also relate to a small but growing literature that studies material conflicts of interest between financial advisers and/or brokers and their clients. Based on archival data, Christoffersen and Musto (2015); Bergstresser et al. (2009); Hackethal et al. (2012); Guerico and Reuter (2014); Hoechle et al. (2018); Fecht et al. (2018); Egan (2019) provide empirical evidence that brokers and advisers direct consumers to high-fee products. Recent theoretical studies focus on the effects of incentive structures and of related policy instruments on financial misbehavior (Inderst and Ottaviani, 2009; Stoughton et al., 2011; Inderst and Ottaviani, 2012a; 2012b; Chang and Szydlowski, 2020). We add to this literature by demonstrating that customers themselves can influence the outcome of financial advice with simple nudges, and by suggesting that unconventional instruments like the banker's oath can complement more traditional regulation.

2. Experimental setup

2.1. Audit study

2.1.1. Scenario and treatments

In order to measure the quality of financial advice and to investigate the behavior of financial advisers, we set up an audit study where advisers are not aware that they are being studied.[3] We sent trained, professional auditors to 201 bank branch offices located all over the Netherlands in 102 cities and villages (see Fig. 1). The offices were operated by two large banks, which, in 2019, had a total of 343 offices in the Netherlands. Hence, this study includes nearly 60% of the two banks' nationwide coverage.[4] The auditors impersonated regular customers, who were seeking financial advice on taking out a car loan. For this, they used a standardized scenario script which described a conflict of interest between the customer and the bank.[5]

Specifically, the auditors indicated that they wanted to buy a car for € 8,000 and were considering to do this with a loan (of the same amount). They also said that they had € 12,000 in savings, but considered to keep that in reserve (without specific plans what to do with that money).[6] Based on pre-checks with online loan requests and pilot visits, we pre-specified simple characteristics that were necessary to be eligible for a car loan of € 8,000, and assigned them to all auditors: they claimed that they were single, without children, that they earned a regular income of € 2,100 net per month (as a temporary worker), and that they had no mortgages or other debts. Advisers were not able to look up more customer information, because the auditors did not have an account at the visited bank, although they indicated to be willing to switch banks. All other characteristics that auditors may have talked about with their advisers were their own characteristics, so that the talk was as natural as possible.

We selected three different treatments that were presented to the advisers (and impersonated by the auditors). In the *control treatment*, CONTROL, auditors presented the above scenario with no further questions or additions.

In the *direct treatment*, DIRECT, the auditors directly mentioned the oath in a statement: "I recently saw in a consumer program (alternatively: read in the newspaper; heard from an acquaintance) that each bank employee has taken the banker's oath." Then they asked about the oath: "What is actually the purpose of the oath?" After the answer of the adviser, the auditor presented the scenario as in CONTROL.

It is possible, however, that the question about the banker's oath simply increases the salience of customer-centered behavior in general, which is the oath's most central element (Loonen and Rutgers, 2017). Therefore, in the *indirect treatment*, INDIRECT, the auditors asked advisers about customers' interests without mentioning the oath directly. First, the auditors remarked: "I have the feeling that my own bank cares more about their own profits than about what is best for their clients." Then they asked as a reminder: "How does your bank protect the interests of their clients?" Thereafter, the auditor proceeded as in Treatment CONTROL. Treatment INDIRECT is an attempt to disentangle the effects linked to customer interests (as the oath's central element) from potentially broader effects that pertain to the underlying norms and values of the oath and its code of conduct.

2.1.2. Outcome variables and expectations

After presenting the above mentioned scenario to the advisers, the auditors asked for advice about the financial possibilities and options. Note that the scenario constituted a situation where customers' interests did not align with advisers' interests if the latter had an incentive to sell financial products. In an environment with significantly higher interest rates for consumer loans than for savings and with additional fees for taking out a loan, the customer-friendly advice would have been to use own savings first before taking up a loan, even after accounting for an emergency buffer. According to a web-based buffer calculator of the Dutch National Institute for Family Finance Information (NIBUD), which is the largest independent information provider on household finance for Dutch consumers, the buffer for this situation is advised to be € 3800 with a minimum of € 3550 (https://bufferberekenaar.nibud.

[3] Generally, the audit study methodology includes a mild form of deception as, for example, resumes of job applicants or loan requests, are not real. However, if the insights gained from an audit study (e.g., consequences for improved institutional design to remedy discrimination and unethical behavior) are considered to outweigh the costs (e.g., time and effort spent to process the applications or the loan requests), then an Institutional Review Board (IRB) can consider this trade-off to be tolerable. After careful deliberation by an ethics commission, our study has gained IRB approval from the University of Innsbruck. We are therefore confident that in our study, the benefits (i.e., insights into the efficiency of the banker's oath) outweigh the costs (i.e., time spent by the financial advisers to consult our auditors).

[4] In this paper, we do not reveal the names of the two banks, because they are not relevant for the interpretation of the results.

[5] All auditors saw the same information material in their preparation for the visits (see the instructions in Appendix B.1).

[6] This scenario is a modified version of a conflict of interest used in Harms (2018). Also see Section 2.1.3 for additional details.

Fig. 1. Location of bank branches visited, audit study: circles indicate $N = 102$ municipalities in the Netherlands where auditors visited in total 201 bank branches.

nl).[7] Hence, there are enough savings to finance the car without the need to take up a loan $(12,000 - 3,800 = 8,200)$.[8] However, if advisers were primarily motivated to increase the loan portfolio of the bank in combination with winning a new customer, they also had the opportunity to push for a product sale and advice the customer to take out a loan.

Directly after the visit, the auditors recorded (on 7-point Likert scales) how strongly the bank employee recommended to take out a loan (L), and, separately, how strongly they recommended to use the savings (S), both ranging from 0 (not at all) to 6 (very strongly). See Section B.2 for the questionnaire. We decided to record the strength of recommendations for S and for L separately, because this allows us to control for levels of advice, in contrast to one relative measure. Pilot visits showed that it was also easier for auditors to record the recommendations separately, particularly in situations where advisers did not strongly recommend any of the two options. Fig. 2 provides a schematic overview of the possible combinations of recommendations for S and L. Our main binary outcome variable is whether a product sale is the priority of the financial advice (LOANPRIO), that is, whether the adviser primarily steers the client towards a loan (LOANPRIO $= 1$ if L $>$ S), thereby serving their own interest, or keeps the client's best interest in mind, by recommending at least an equal amount of own savings (LOANPRIO $= 0$ if L $\leq$ S). If a direct nudge is necessary to remind financial advisers of their oath, we would expect less advice with product sales as a priority in Treatment DIRECT than in CONTROL, that is, a lower likelihood for observations below the dashed diagonal line in Fig. 2, as indicated by the arrow from LOANPRIO $= 1$ to LOANPRIO $= 0$. Analogously, if advisers simply need to be reminded of customers' interests, we expect the same for the indirect nudge in Treatment INDIRECT.

Hypothesis 1. LOANPRIO $= 1$ is less likely in Treatment DIRECT than in Treatment CONTROL.

Hypothesis 2. LOANPRIO $= 1$ is less likely in Treatment INDIRECT than in Treatment CONTROL.

As a second outcome measure, we compute the net strength of advice in the bank's interest: LOANSTRENGTH $=$ L $-$ S.[9] The rationale behind this measure is that it is more finely grained than the dichotomous variable LOANPRIO, but this comes at the cost of also capturing more noise. Fig. 2 illustrates it with two hypothetical recommendations: one advising to use loans with a net strength of LOANSTRENGTH $= -1$ and another with LOANSTRENGTH $= 3$. If a nudge is effective we expect that the net strength of advice for using loans is smaller in DIRECT (or in INDIRECT) than in CONTROL, as indicated by the grey areas in Fig. 2. For example, if position 5,2 is the advice in favor of loans in the control treatment, then, after a nudge, we would expect a shift above the local, solid 45° line (intercepting 5,2) with LOANSTRENGTH < 3. Note that this includes shifts where the treated advice still prioritizes loans (i.e., the advice stays below the dashed diagonal line intercepting 0,0) and where the advice for taking out loans may even increase. For example, if a nudge increases the strength of advice for loans from 5 to 6, but also the advice for using own savings from 2 to 4, then the net advice for using loans has decreased from LOANSTRENGTH $=$ 3 (5,2) to LOANSTRENGTH $= 2$ (6,4).

Hypothesis 3. LOANSTRENGTH is smaller in Treatment DIRECT than in Treatment CONTROL.

Hypothesis 4. LOANSTRENGTH is smaller in Treatment INDIRECT than in Treatment CONTROL.

As explained in Section 2.1.1, we administer INDIRECT as a comparison treatment to test whether a direct reminder of the banker's oath (DIRECT) merely increases the salience of customers' interests or triggers a stronger effect, arguably by (re-)activating a whole set of norms and values. In case of the latter, and without formulating a separate hypothesis on this, we expect that support for Hypotheses 1 and 3 is stronger than for Hypotheses 2 and 4.

2.1.3. Implementation of the audit

Before the implementation of the audit study, we piloted two different scenarios with 20 observations each, from April 16, 2019, to May, 2, 2019, to find a suitable scenario and to test scenario scripts as well as questionnaires. The visit logs of the pilot study indicated that our selected scenario was able to (i) generate sufficient advice that pushed product sales, so that there was enough potential for a nudge to have an effect, and (ii) reliably generate data in one visit without the risk that the adviser asked for a second meeting. Note that due to changes in the scenario scripts and questionnaires the pilot data is not included in the analyses.

The audit field data for this study was collected from August 15, 2019 to March 3, 2020 with 201 bank branch visits. By sheer coincidence, our data collection phase ended about two weeks before public mobility in the Netherlands was severely restricted due

[7] NIBUD is a non-profit foundation, founded in 1979, with the goal to prevent money problems of consumers. Almost twenty percent of NIBUD's activities are financed by the Dutch Ministry of Social Affairs and Employment and the Dutch foundation BKR (Credit Registration Office). The NIBUD buffer calculator was first launched in 2008, is regularly updated, and well-known in the Netherlands. In fact, some of the advisers referred to NIBUD's buffer calculator during the visit themselves. 86% of Dutch consumers know NIBUD and 35% visited NIBUD's website, based on an Imago survey in 2019 (https://www.nibud.nl/wp-content/uploads/Nibud-Factsheet-Bereik-en-Imago-2020.pdf). For details on how the buffer is computed see https://www.nibud.nl/wp-content/uploads/A_reference_buffer_for_households.pdf.

[8] Note that the buffer calculation and data collection was completed before the onset of the COVID-19 pandemic where less extreme financial situations were expected.

[9] Note that LOANPRIO $= 1$ if LOANSTRENGTH > 0 and LOANPRIO $= 0$ if LOANSTRENGTH ≤ 0.

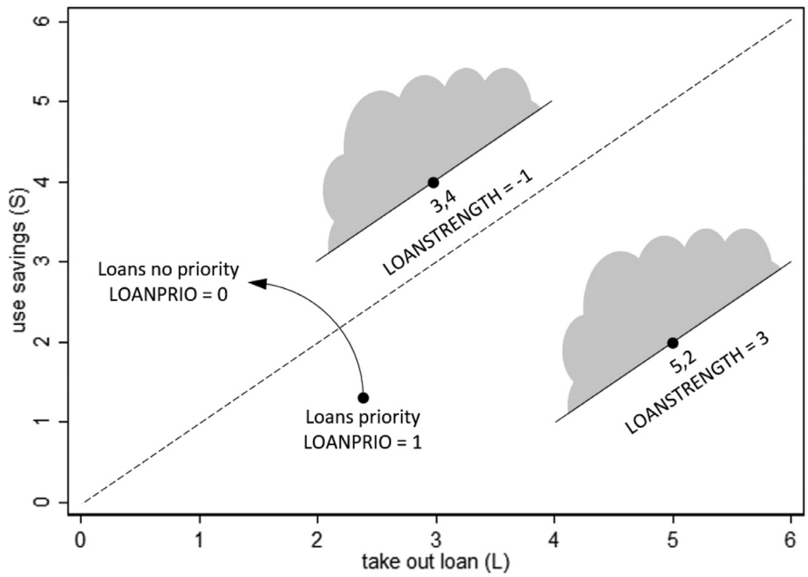

Fig. 2. Outcome variables and expected effects: LOANPRIO = 1 indicates a stronger advice for loans than savings: L > S. LOANSTRENGTH measures the net strength of the advice for loans: L − S. If nudges are effective, (i) LOANPRIO = 1 is less likely in Treatment DIRECT and in Treatment INDIRECT than in Treatment CONTROL, as indicated by the arrow; and (ii) $\text{LOANSTRENGTH}^{dir} < \text{LOANSTRENGTH}^{con}$ as well as $\text{LOANSTRENGTH}^{ind} < \text{LOANSTRENGTH}^{con}$, as indicated by the grey areas.

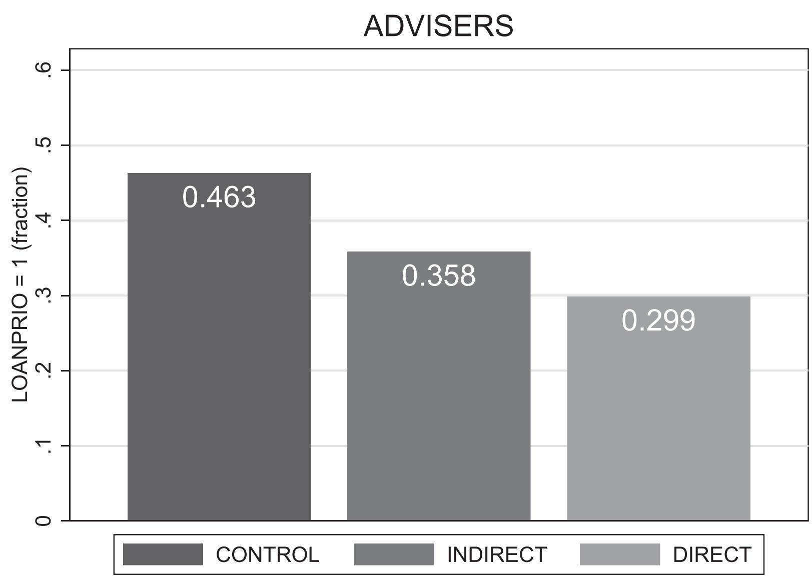

Fig. 3. Fraction of advice to primarily take out a loan, audit study: CONTROL refers to the control treatment. In Treatment DIRECT, auditors administered a nudge that directly reminded financial advisers of the oath they took. In Treatment INDIRECT, auditors implemented an indirect nudge that reminded advisers of the customer's interests. LOANPRIO indicates whether the adviser primarily steers the client towards a loan (LOANPRIO = 1 if L > S).

to the outbreak of COVID-19.[10] We were thus able to collect all observations as originally planned, with more than 80%, 90%, and 98% of observations before January, February, and March 2020, respectively. To control for possible unobserved confounds related to COVID-19 we nevertheless include time trend controls in our analyses.

[10] In the Netherlands, the first positively tested case of COVID-19 was reported on February 27, with the first casualty on March 6. On March 12, the Dutch government banned gatherings of more than 100 people and advised everyone to work from home where possible. On March 15, universities, schools, child care centers, bars, restaurants, hairdressers, sports clubs and other facilities were closed. Until March 26, several additional restrictions were announced, including the ban of public gatherings of more than two people and the requirement for 1.5 m inter-personal distance in public.

To implement the visits, we hired a professional audit firm that specializes in identifying and training auditors. We worked very closely with the audit firm to develop the scenario scripts, to select the two banks with sufficient bank branches across the country, and to set up the schedule of visits. The audit firm provided the logistics of monitoring and implementing the scheduling of visits, finding and compensating auditors, and providing the mobile application for the exit questionnaires. We randomly assigned 201 advisers (bank branches) from both bank networks to each of the three treatments. Each auditor was assigned to each of the three treatments at least once to control for auditor fixed effects. Auditors did not know the purpose of the study and the sequence of the assignments was random and balanced across auditors. Most auditors (42 out of 51) completed only one sequence of three visits. For auditors with more than three visits, each sequence needed to be completed in a separate month in order to maintain bal-

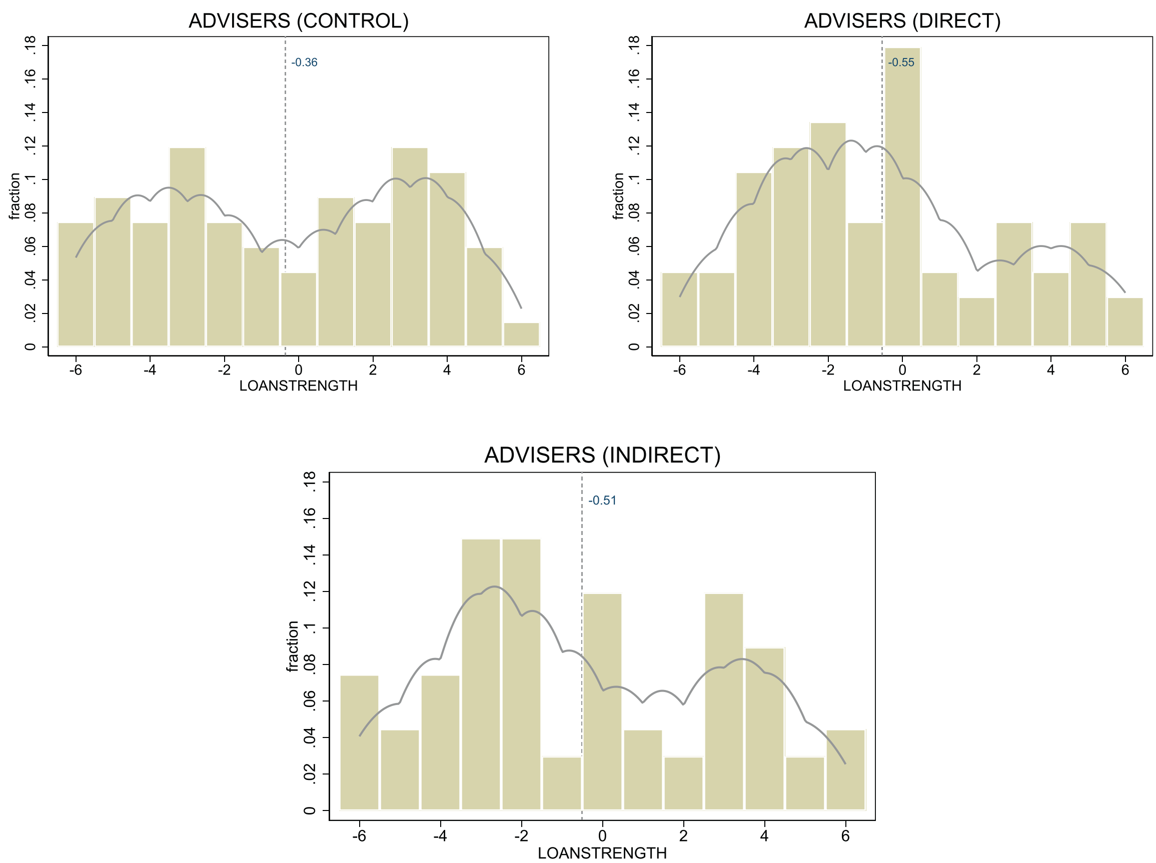

Fig. 4. Net strength of advice for loans (LOANSTRENGTH), audit study: net strength of advice (LOANSTRENGTH) for loans is calculated as LOANSTRENGTH = L − S. Histograms per treatment with Epanechnikov kernel distributions are displayed. Vertical dashed lines show averages of LOANSTRENGTH.

ance across treatments and prevent clustering over time.[11] Hence, we follow a matched pair design, where every strata of auditor and month contains just three advisers, each of which is randomly assigned to one of three treatments (see, e.g., Gerber and Green, 2012). Moreover, per treatment, we tried to keep the proportion of bank network affiliations of the advisers as similar as practically possible.[12] Auditors were paid on a per visit basis. No adviser (bank branch) was visited more than once.

The procedure of the bank visits was as follows. The auditor visited a specific bank branch that was centrally assigned to him/her by the audit firm with a special scheduling software. After entering the bank, the auditor asked for an employee who can provide ad-hoc financial advice on loans.[13] The auditor was trained to impersonate a customer who was clearly interested in taking out a loan without ultimately buying the financial product. Directly after visiting the bank, the auditor recorded the received advice in an exit questionnaire, together with some other variables of interest, such as timestamps, the (estimated) age and gender of the adviser, and the strength of the reaction of the adviser in answering the question on the nudge. For the full exit questionnaire, please see Appendix B.2. All responses were recorded with a mobile application and uploaded, together with a picture of the bank branch, to the audit firm, which forwarded the data to the research team (after replacing the auditor identity with an anonymous identifier).

[11] Five auditors completed two sequences (six visits) and four completed three to five sequences (nine to 15 visits).

[12] A perfect block randomization per bank was operationally not possible, because of the distances some auditors would have been required to travel.

[13] If the adviser was available but did not contact the auditor within 10 minutes, the auditor was allowed to actively approach the adviser.

2.2. Prediction task

Randomized controlled trials have many advantages, such as causal inference and a higher external validity than, for example, lab experiments, but they are often operationally limited with regard to the number of observations. Hence, there is a realistic chance for a null result, which gives rise to two potential problems. First, there is a risk of reporting false negatives (Type II errors). Power tests can alleviate these concerns, but crucially depend on the effect size that we can realistically expect to find. Second, once the results are known, researchers and peers suffer from hindsight bias, which makes it difficult to appreciate the novelty and contribution of the study (DellaVigna et al., 2019). Despite preregistration, hindsight bias may particularly apply to studies with null results, which are rarely published even when they answer important questions with rigorous methods (Franco et al., 2014). To mitigate these problems, we follow DellaVigna et al. (2019) and collected the beliefs (priors) about treatment effects of relevant groups before the results were known. This provides us with both an expected treatment effect for power tests and an ex ante prediction that can be compared with the audit results, making the latter more informative. Specifically, before the audit study was completed, we sent out two online surveys with an incentivized prediction task that elicited the expected outcome of the audit study.

In the online survey EXPERTS, we administered a prediction task to 122 Dutch policy experts from the Dutch Central Bank, the Dutch Authority for Financial Markets (AFM), the Netherlands Bureau for Economic Policy Analysis (CPB), the Dutch Ministry of Finance, the Dutch Ministry of Economics, and the “Behavioural Insights Netwerk Nederland” (BIN NL), which is an alliance of all Dutch ministries for the application of behavioral knowledge within central government. The survey was distributed from February 11, 2020 to March 11, 2020 and took the median respondent 4.6 minutes to complete.[14] The survey started with a description of our audit experiment, followed by a task to predict the average Likert-scale-answers for the two variables S and L as recorded by the auditors. Every respondent first predicted the outcome in Treatment CONTROL, followed by, in a randomized order, Treatments DIRECT and INDIRECT. Respondents were informed that, if a randomly drawn prediction was within $+/-0.2$ points around the real average of the audit study, they received € 25, following Cohn et al. (2014).[15] The survey ended with a few questions on gender, age, job function, and job experience in years as well as in comparison to colleagues. Please see Appendix B.3 for the full instructions.

In the second online survey CUSTOMERS, we sent the same prediction task to a representative sample of the Dutch working population ($N = 502$), stratified by gender, age, education, and region.[16] We chose this sample as an approximation for Dutch bank customers who potentially apply for a car loan. The survey was distributed from February 18, to 21, 2020 and took the median respondent 3.8 minutes to complete. The main screens of the survey and also the incentives were identical to Survey EXPERTS. In the exit questionnaire, we measured financial literacy as in van Rooij et al. (2011) – with slightly modified questions to impede online lookup – and added a question on personal experience with taking out bank loans. Please see Appendix B.4 for the full instructions.

2.3. Summary statistics and randomization check

Table 1 provides an overview over all samples from the audit study (AUDITORS and ADVISERS) and the two online prediction surveys (EXPERTS and CUSTOMERS). In the audit study, the average length of a bank visit was 19 minutes, including an average 10 minutes talk with the financial adviser. Importantly, 'reaction to DIRECT Q' (... INDIRECT Q) is the strength of the response of the adviser to the direct (indirect) nudging questions, measured on a Likert-scale from 0 (not at all) to 6 (very much). With an average score of 4 to the direct question about the oath (and with a 2.9 to the indirect question), advisers clearly reacted to the respective nudge of the customers. We are thus confident that the nudge was recognized by the adviser and that the treatment was administered successfully.

The EXPERTS in the online prediction task have an average work experience of 6.6 years, consider themselves to be close to average (2.6/5) in their self-assessed work experience in projects that are related to our audit study, and mostly have job functions that are related to regulation, policy, and supervision (72%), and/or research and analyses (39%).[17] Given this profile, we are confident that the experts are sufficiently knowledgeable to make informed predictions about the audit study. The potential CUSTOMERS are financially quite literate with, on average, 2.13 correct answers. Moreover, the majority of respondents (59%) has prior experience in taking out a personal loan (for example, a consumer loan or a mortgage).[18] Given these profiles, we are confident that both the experts and the customers are sufficiently knowledgeable to provide, on average, informed (i.e., non-random) predictions about the audit study.

Finally, we test whether the randomization of bank branches and advisers across treatments has been successful. For this, we run a multinomial logit with the adviser data from the audit study (henceforth, ADVISERS) and a categorical variable indicating the three treatments as the dependent variable (CONTROL is the baseline). As reported in Table A2 in the Appendix, no potentially confounding variable, such as bank affiliation, adviser demographics, or the timing of the visit, predicts any treatment affiliation.[19] We can therefore conclude that the allocation of advisers to treatments was indeed random with no interference of other factors.

3. Results

The main focus lies on testing the hypotheses with the audit study, i.e., with the sample ADVISERS. We organize the results of the audit study along the main outcome variables LOANPRIO and LOANSTRENGTH. Subsequently, we analyze the predictions of EXPERTS and CUSTOMERS, which allow for supplementary comparisons and power tests.

3.1. Audit study (*advisers*)

3.1.1. Advice that prioritizes product sales (*loanprio*)

Result 1. Without any intervention, nearly half of all financial advisers (46.3%) primarily push loans in their recommendations. Direct nudges that remind financial advisers of their oath significantly decrease the likelihood that recommendations prioritize product sales (loans).

Support: Fig. 3 reports the fraction of financial advice that prioritizes loans across treatments: the fraction of LOANPRIO $= 1$ in Treatment CONTROL, where no nudge was applied, is 0.463 (46.3%). Figure A1 in the Appendix provides more detail by displaying the financial advice in each of the 67 visits in Treatment CONTROL of the audit study.[20] In particular, we observe from Fig. 3 that the fraction of LOANPRIO $= 1$ decreases from 0.463 (46.3%) in Treatment CONTROL to 0.299 (29.9%) in Treatment DIRECT for ADVISERS. This is a substantial drop in the prioritization of product sales of more than 16 percentage points.

As main analysis, we run a logistic regression with treatment dummies on LOANPRIO as the dependent variable. Models 1 and 2 in Table 2 report the results (odds ratios) with clustered standard errors per auditor. Model 1 regresses the treatment dummies on LOANPRIO. Model 2 accounts for a number of control variables, including the overall 'level' of the strength of financial advice (S + L), the bank affiliation of the advisor, the month and the order of the visit as trend variables, and some adviser and auditor demographics. We find that the control variable 'level' is correlated with LOANPRIO with an odds ratio below one, indicating that the prioritization of product sales is more likely in meetings where finan-

[14] Although the survey was also administered after the end of the audit study (March 3), the results of the audit study were not publicly known until the end of the survey period. Our survey results do not change significantly if we restrict our sample to the 110 experts who answered the survey by March 3, 2020.

[15] Also, for transparency reason and as an additional incentive, they were promised early access to the relevant findings of the audit study.

[16] For this we used the services of the market research firm Dynata (https://www.dynata.com/), which is able to pay out decision-dependent incentives to respondents and to provide ex post feedback via email.

[17] Percentages do not add up to 100, because of multiple job functions per expert. All results reported in this study are robust to a reduction of the sample to (i) experts with job functions in regulation, policy, and supervision, and (ii) experts who consider themselves at least as experienced in projects that are related to our audit study than the average colleague in their organization (experience relative≥ 3).

[18] All results reported in this study are robust to a reduction of the sample to (i) customers with a financial literacy score of 3/3, and (ii) customers who have previously taken out a personal loan.

[19] Note that all auditor characteristics are balanced across treatments by design (see Section 2.1.3) and are therefore not included in the estimation.

[20] Recall that any advice below the 45° diagonal (intercepting 0,0) is primarily in favor of taking out a loan (LOANPRIO $= 1$). The scatterplot highlights the observations below the diagonal with red circles.

Table 1
Descriptive statistics for all samples: 'female' is a categorical variable (1, else 0) for the gender of the respondents (rsp) and of the advisers (ADVISERS); 'age rsp' is in years for EXPERTS and CUSTOMERS. In the audit study, the age of the AUDITORS and the (estimated) age of the ADVISERS is recorded in age brackets (21–30, 31–40, 41–50, and 51–59 or >51, respectively), which we transform into rounded midpoints per bracket (26, 36, 46, 56); 'bank' is a dummy recording which of the two national banks the visited branch belongs to; 'nr of visits' is the amount of visits per auditor (one per branch); the length of the visit (including waiting times) and the length of the talk with the adviser is in minutes; 'reaction to DIRECT Q' (... INDIRECT Q) is the strength of the response of the adviser to the direct (indirect) nudging questions, measured on a Likert-scale from 0 (not at all) to 6 (very much); 'experience in yrs' is the number of years of work experience related to regulation and/or policy work; 'experience relative' is the self-assessed work experience in projects/topics that are related to the audit study, compared to the average colleague in the organization; 'work in reg/pol' equals 1 if any one of the expert's job function is regulation and/or policy work and/or supervision; analogously, 'work in res/analyses' equals 1 for research and/or analyses; 'loan experience (Y/N)' equals 1 if the customer has experience with taking out a personal loan (for example, a consumer loan or a mortgage); 'financial literacy' records the number of correct answers to the three (modified) financial literacy questions of van Rooij et al. (2011).

	mean	sd	min	max	N
AUDIT STUDY					
AUDITORS					
female rsp	0.57	0.5	0	1	51
age rsp	47.96	9.6	26	56	51
no of visits	3.94	2.44	3	15	51
ADVISERS					
female adv	0.46	0.5	0	1	201
age adv	37.24	8.77	26	56	201
bank	0.60	0.49	0	1	201
length visit (min)	18.62	12.31	2	89	201
length talk (min)	9.61	5.68	2	41	201
reaction to DIRECT Q	4.03	1.63	0	6	67
reaction to INDIRECT Q	2.94	1.67	0	6	67
SURVEYEXPERTS					
female rsp	0.4	0.49	0	1	122
age rsp	37.95	10.46	24	65	122
experience in yrs	6.57	6.5	0	33	122
experience relative	2.61	1.17	1	5	122
work in reg/policy	0.72	0.45	0	1	122
work in res/analyses	0.39	0.49	0	1	122
SURVEY CUSTOMERS					
female rsp	0.5	0.5	0	1	502
age rsp	43.6	13.17	19	66	502
loan experience (Y/N)	0.59	0.49	0	1	502
financial literacy	2.13	0.98	0	3	502

Table 2
Estimations on LOANPRIO, audit study: Logistic regressions with LOANPRIO as the dependent variable and clustered standard errors per auditor. 'DIRECT' and 'INDIRECT' are treatment dummies. 'permute p' (table bottom) reports the p-values of the corresponding treatment dummy coefficients, obtained from permutation tests with 1000 random draws (accounting for respondent strata). 'level' is S + L. The month of the visit and the position of the visit per auditor and month (pos=1,2,3) are included as trend variables. Remaining variables are defined in the notes of Table 1. The table reports odds ratios with z-values in parenthesis. *0.05 **0.01 ***0.001 denote levels of statistical significance.

	(1)	(2)
DIRECT	0.494*	0.482*
	(−2.29)	(−2.18)
INDIRECT	0.648	0.626
	(−1.59)	(−1.60)
level		0.672*
		(−2.52)
bank		1.054
		(0.16)
pos		0.686
		(−1.93)
month		0.931
		(−0.70)
female adv		1.135
		(0.41)
age adv		0.996
		(−0.22)
female rsp		0.697
		(−0.78)
age rsp		0.967
		(−1.29)
constant	0.861	2.59e+24
	(−0.55)	(0.76)
permute p DIRECT	0.036	0.042
permute p INDIRECT	0.191	0.145
Prob> χ^2	0.062	0.089
N	201	201

cial advice is generally less strong.[21] In both models, the dummy for the Treatment DIRECT is statistically significant with odds ratios 0.494 and 0.482. The marginal effect of DIRECT is -.162 in Model 1 and -.151 in Model 2.[22] This means that, after mentioning the oath (DIRECT = 1), advice that pushes loans (LOANPRIO = 1) is 15.1 to 16.2 percentage points less likely than in the control treatment, in support of Hypothesis 1.

In addition, we also run permutation tests of all models with 1000 random draws each, accounting for auditor strata.[23] Permutation tests simulate the Null with random treatment assignments and record how often the simulated coefficient of the treatment variable is greater than the observed coefficient (or odds ratio, as reported in Models 1 and 2). The less often a random treatment allocation beats the observed treatment effect, the more likely it is that the actual treatment allocation caused the observed effect.[24] At the bottom of Table 2 we report, for both treatment dummies, the permutation p-values, which indicate how likely the observed treatment coefficients are an outcome of a random allocation. The permutation p-values for DIRECT are clearly below the 5% level of significance, with $p = 0.036$ in Model 1 and $p = 0.042$ in Model 2. Hence, also the results of the permutation tests fully support Hypothesis 1.

Finally, as a robustness check, we compute McNemar's tests of 2×2 contingency distributions of LOANPRIO across treatments, as shown in Table A1 in the Appendix. All McNemar's tests reject the Null that the discordant proportions of LOANPRIO across Treatments DIRECT and CONTROL are equal (with $\chi^2 = 4.17$ and $p = 0.041$). Again, the result of the robustness check is in support of Hypothesis 1, indicating that a direct nudge, referring to the oath, significantly decreases the likelihood that recommendations prioritize product sales.

Result 2. Indirect nudges that remind financial advisers of customers' interests do not significantly affect the likelihood that recommendations prioritize product sales.

[21] All results reported in Table 2 are robust to the exclusion of the variable 'level'. We do not find statistically significant interaction effects between the treatment dummies and (i) advisor age, (ii) advisor gender, (iii) auditor age, and (iv) auditor gender.

[22] For the marginal effects analyses please see the online supplementary data/code. The corresponding values for INDIRECT are -.099 (Model 1) and -.097 (Model 2), but statistically insignificant.

[23] We use the user programmed command ritest in Stata, described in Heß (2017).

[24] In contrast to classical inference, permutation tests and randomization inference (Fisher, 1935) do not require large samples drawn from infinite populations, relying on asymptotic properties of estimators. Permutation tests are therefore often the preferred methodology for experiments (Imbens and Rubin, 2015).

Support: As Fig. 3 shows, the unconditional fraction of LOANPRIO = 1 decreases from 0.463 in Treatment CONTROL to 0.358 in Treatment INDIRECT. In the regression analysis, reported in Table 2, the coefficients of the dummy for Treatment INDIRECT in Models 1 and 2 (with LOANPRIO as dependent) are statistically insignificant. The same applies to the p-values of the corresponding permutation tests for the coefficients of INDIRECT reported at the bottom of the table ($p = 0.191$ for Model 1 and $p = 0.145$ for Model 2) and to McNemar's tests, reported in Table A1 in the Appendix.

Overall, we conclude that we do not find support for Hypothesis 2. Importantly, the fact that Hypothesis 1 is clearly supported (Result 1) while Hypothesis 2 is not, provides evidence for the notion that a direct nudge that reminds advisers of the banker's oath speaks to a different mechanism than merely increasing the salience of customers' interests.

3.1.2. Net strength of advice to use loans (loanstrength)

Result 3. Without any intervention, financial advice follows a bimodal distribution, either leaning toward using own savings, or toward taking out a loan. Nudges do not decrease net advice to take out loans, but they increase the frequency of neutral advice.

Support: Fig. 4 displays distributions of the net strength of financial advice (LOANSTRENGTH) in the audit study across treatments: the top panel with observations from Treatment CONTROL shows a quite symmetric bimodal distribution with modes at $\text{LOANSTRENGTH}^{con} = -3$ and $\text{LOANSTRENGTH}^{con} = 3$, each with a fraction of 11.9% of all financial advice. As a comparison, neutral advice, which neither favors loans nor savings ($\text{LOANSTRENGTH}^{con} = 0$), makes up only 4.5% of all observations.[25] As reported in Table A3, $\text{LOANSTRENGTH}^{con}$ is statistically not different from zero (Wilcoxon signed-rank test with $z = -0.83$, $p = 0.41$). When focusing on treatment differences in LOANSTRENGTH, Table A3 in the Appendix reveals that none of the treatments DIRECT and INDIRECT have a statistically significant effect on S, L, and LOANSTRENGTH (Table A3 displays the z- and p-values of Wilcoxon signed-rank tests).

Additional tests also show that there is little evidence for treatment differences regarding variable LOANSTRENGTH. Table 3 reports the results of panel regressions of treatment dummies on LOANSTRENGTH with absorbed fixed effects at the auditor level. The bottom of the table displays the p-values of the corresponding permutation tests. In line with the paired two-sample tests in Table A3 in the Appendix, the coefficients of DIRECT and INDIRECT are not significant, and neither are the p-values of the permutation tests.[26]

One reason for this finding is that nudges increase neutral financial advice, as shown in Fig. 4. In Treatment CONTROL only 4.5% of all advice was neutral ($\text{LOANSTRENGTH}^{con} = 0$) with two modes at $\text{LOANSTRENGTH}^{con} = \pm 3$. After the direct nudge in Treatment DIRECT, however, $\text{LOANSTRENGTH}^{dir} = 0$ is the mode with 17.9% of all observations and the mass of the distribution shifted into non-positive territory. The latter also applies to the indirect nudge, INDIRECT, with modes at values below zero ($\text{LOANSTRENGTH}^{ind} = -2$ and $\text{LOANSTRENGTH}^{ind} = -3$), and, again, $\text{LOANSTRENGTH}^{ind} = 0$ is more frequent (11.9%) than $\text{LOANSTRENGTH}^{con} = 0$ (4.5%). To test the effect of nudges on neutral advice, we run logistic regressions and permutation tests with a dummy variable for neutral advice (LOANSTRENGTH = 0) as the dependent variable. As reported in Table 4, the direct nudge, asking advisers about their oath in Treatment DIRECT, makes neutral advice 4.7 to 5.5 times more likely, depending on the econometric model. Treatment INDIRECT turns out to be mostly insignificant though close to the 5% level both in the logistic regressions and in the permutation tests.

Hence, overall, we conclude that the data from the audit study do not provide support for Hypotheses 3 and 4. Nudges primarily reduce the bimodality of the distribution in Treatment CONTROL, by adding more neutral advice, but have little effect on means and medians.

Table 3
Estimations on LOANSTRENGTH, audit study: Panel regressions with LOANSTRENGTH as the dependent variable and absorbed auditor fixed effects (correspondingly, separate auditor controls are dropped). 'DIRECT' and 'INDIRECT' are treatment dummies. 'permute p' (table bottom) reports the p-values of the corresponding treatment dummy coefficients, obtained from permutation tests with 1000 random draws (accounting for respondent strata). 'level' is the overall strength of financial advice, i.e., S + L. 'bank' is a dummy for one of two bank networks the visited office belongs to. The month of the visit and the position of the visit per auditor and month (pos=1,2,3) are included as trend variables. 'female' is a categorical variable (1, else 0) for the gender of the advisers (adv). 'age' of the ADVISERS is the rounded midpoint per estimated age bracket (26, 36, 46, 56). The table reports coefficients with t-values in parenthesis (robust). *0.05 **0.01 ***0.001 denote levels of statistical significance.

	(1)	(2)
DIRECT	−0.194	−0.132
	(−0.36)	(−0.25)
INDIRECT	−0.149	−0.145
	(−0.29)	(−0.29)
level		−0.424*
		(−2.13)
bank		0.509
		(1.08)
pos		−0.653*
		(−2.58)
month		−0.263
		(−1.27)
female adv		0.482
		(0.89)
age adv		−0.019
		(−0.63)
constant	−0.358	191.979
	(−0.92)	(1.29)
permute p DIRECT	0.715	0.787
permute p INDIRECT	0.780	0.770
Prob>F	0.931	0.114
N	201	201

3.2. Predictions of experts and customers

Result 4. Experts predict that, without any intervention, financial advisers prioritize product sales (loans) more often than observed in the field. Moreover, experts predict that, both, direct and indirect nudges significantly reduce the prioritization of product sales. In the field, this effect applies only for direct nudges, which is correctly anticipated by customers.

Support: Fig. 5 replicates Fig. 3 from the audit study and reports the fraction of LOANPRIO = 1 in the samples EXPERTS ($N = 122$) and CUSTOMERS ($N = 502$). In Treatment CONTROL (without any intervention), EXPERTS predict that product sales will be prioritized by financial advisors in 63.1% (0.631) of all cases. This is higher than the corresponding value of 46.3% (0.463) that we find in the audit study (see Fig. 3). A two-sided Fisher's χ^2 exact test

[25] A dip test (Hartigan and Hartigan, 1985) rejects the Null of unimodality with $p = 0.033$.

[26] The marginal effects of DIRECT (INDIRECT) are -.151 (-.104) in Model 1 and -.155 (-.109) in Model 2; all statistically insignificant. For the marginal effects analyses please see the online supplementary data/code.

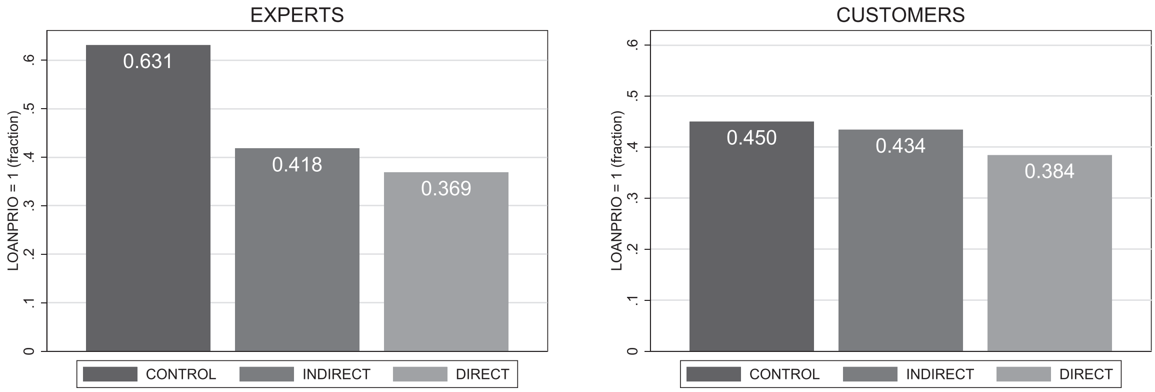

Fig. 5. Fraction of advice to primarily take out a loan, surveys: CONTROL refers to participants' predictions for the control treatment. In DIRECT, participants predict the fraction to take out a loan based on the audit results in Treatment DIRECT. In INDIRECT, participants predict the outcome of Treatment INDIRECT. If LOANPRIO = 1, participants predicted advice to primarily take out a loan.

Table 4
Estimations on neutral advice, audit study: Logistic regressions with a dummy for neutral advice (i.e., value of 1, if LOANSTRENGTH = 0, zero otherwise) as the dependent variable and clustered standard errors per auditor. 'DIRECT' and 'INDIRECT' are treatment dummies. 'permute p' (table bottom) reports the p-values of the corresponding treatment dummy coefficients, obtained from permutation tests with 1000 random draws (accounting for respondent strata). 'level' is the overall strength of financial advice, i.e., S + L. 'bank' is a dummy for one of two bank networks the visited office belongs to. The month of the visit and the position of the visit per auditor and month (pos=1,2,3) are included as trend variables. 'female' is a categorical variable (1, else 0) for the gender of the advisers (adv) and auditors (rsp). 'age' of the ADVISERS is the rounded midpoint per estimated age bracket (26, 36, 46, 56). 'age rsp' is the age of the AUDITORS in years. The table reports odds ratios with z-values in parenthesis. *0.05 **0.01 ***0.001 denote levels of statistical significance.

	(1)	(2)
DIRECT	4.655*	5.472*
	(2.36)	(2.43)
INDIRECT	2.893	3.323
	(1.55)	(1.66)
level		0.800
		(−1.42)
bank		2.194
		(1.40)
pos		0.988
		(−0.04)
month		1.068
		(0.54)
female adv		0.999
		(−0.00)
age adv		0.993
		(−0.26)
female rsp		0.299*
		(−2.05)
age rsp		1.039
		(1.41)
constant	0.047***	0.000
	(-5.04)	(−0.58)
permute p DIRECT	0.014	0.012
permute p INDIRECT	0.047	0.054
Prob> χ^2	0.060	0.083
N	201	201

shows that this difference (of 16.8 percentage points) is statistically significant ($p = 0.031, N = 67 + 122 = 189$). In contrast, the prediction of the sample CUSTOMERS (0.45 in CONTROL) is statistically not different from the corresponding fraction in the audit study ($p = 0.474, N = 67 + 502 = 569$).[27] Thus, it seems that the expectations of CUSTOMERS about the quality of financial advisers' recommendations in the Treatment CONTROL are quite realistic while EXPERTS have an overly pessimistic view in the control treatment (CONTROL).

According to Result 1 from the audit study, direct nudges significantly decrease the prioritization of product sales. This finding is correctly anticipated by both experts and potential customers. As shown in Fig. 5, the fractions of LOANPRIO = 1 decrease from 0.631 in Treatment CONTROL to 0.369 for EXPERTS, and from 0.450 to 0.384 for CUSTOMERS. The test statistics are reported in Table 5, which replicates Table A1 from the audit study. McNemar's test rejects the Null for the discordant proportions of LOANPRIO across treatments DIRECT and CONTROL with $\chi^2 = 28.44$ and $p < 0.001$ (with $\chi^2 = 8.44$ and $p = 0.004$) for EXPERTS (for CUSTOMERS). Hence, with regard to LOANPRIO both samples correctly anticipate the effectiveness of mentioning the oath as a direct nudge intervention.

The sample EXPERTS, however, also predicts that indirect nudges are effective (with $\chi^2 = 21.13$ and $p < 0.001$ for Treatment INDIRECT in Table 5), which is in contrast to the Null result in the audit study (Result 2). Interestingly, the sample CUSTOMERS does not expect any effects of the indirect nudge (with $\chi^2 = 0.46$ and $p = 0.496$).

Result 5. Experts predict that nudges significantly shift the net strength of advice away from loans (although these effects are not observed in the field). Customers also expect treatment effects, but they are statistically not different from the observations in the field.

Table 6 replicates the statistics for LOANSTRENGTH in Table A3 (audit study). As we can see from the Wilcoxon signed-rank tests (signed-rank z and p values), both samples, EXPERTS and CUSTOMERS, expect clear treatment effects of DIRECT as well as INDIRECT on LOANSTRENGTH (with $p = 0.00$).[28] To test whether these predicted effects are actually different from the corresponding effects in the audit study, we run a diff-in-diff

[27] All other fractions reported in Fig. 5 (pertaining to Treatments DIRECT and INDIRECT) are statistically not different from the corresponding fractions in the audit study (see Fig. 3).

[28] Figure A2 in the Appendix complements Tables A3 and Table 6 by displaying the distributions of LOANSTRENGTH across samples and treatments.

Table 5
Matched case-control distributions of LOANPRIO, surveys: 2x2 contingency tables per survey sample (EXPERTS, CUSTOMERS) with the outcome frequencies of LOANPRIO in treatments CONTROL V DIRECT and CONTROL V INDIRECT. Observations are matched per respondent. 'Discordant proportions' refer to switches in LOANPRIO across treatments. χ^2 and Prob> χ^2 are from a McNemar's test of the Null that the discordant proportions are equal. *0.05, **0.01, ***0.001 denote levels of statistical significance.

EXPERTS							
		DIRECT			INDIRECT		
		LOANPRIO = 0	LOANPRIO = 1	Sum	LOANPRIO = 0	LOANPRIO = 1	Sum
	LOANPRIO = 0	43	2	45	42	3	45
CONTROL	LOANPRIO = 1	34	43	77	29	48	77
	Sum	77	45	122	71	51	122
Discordant proportions		0.279	0.016*		0.238	0.025*	
	χ^2	28.44			21.13		
McNemar	Prob> χ^2	0.000***			0.000***		
CUSTOMERS							
		DIRECT			INDIRECT		
		LOANPRIO = 0	LOANPRIO = 1	Sum	LOANPRIO = 0	LOANPRIO = 1	Sum
	LOANPRIO = 0	228	48	276	211	65	276
CONTROL	LOANPRIO = 1	81	145	226	73	153	226
	Sum	309	193	502	284	218	502
Discordant proportions		0.161	0.096		0.145	0.129	
	χ^2	8.44			0.46		
McNemar	Prob> χ^2	0.004**			0.496		

Table 6
Strength of advice (LOANSTRENGTH), prediction task: reports means, standard deviations (sd) and Wilcoxon test statistics for LOANSTRENGTH per treatment (CONTROL, DIRECT, INDIRECT) and sample (EXPERTS and CUSTOMERS, with $N = 122$ and $N = 502$ observations per cell, respectively). LOANSTRENGTH is L − S per respondent. The signed-rank z- and p-values (without parentheses) are obtained from Wilcoxon tests of the matched pairs $\text{LOANSTRENGTH}^{con} = \text{LOANSTRENGTH}^{dir}$ and $\text{LOANSTRENGTH}^{con} = \text{LOANSTRENGTH}^{ind}$ within sample. The z- and p-values in parentheses refer to Wilcoxon tests that $\text{LOANSTRENGTH}^{con}$ is equal to zero. The diff-in-diff p-value in the bottom row refers to Wilcoxon rank-sum tests that the treatment effects in samples EXPERTS and CUSTOMERS are equal to the corresponding treatment effects in the audit study (sample ADVISERS); that is whether $\text{LOANSTRENGTH}^{dir} - \text{LOANSTRENGTH}^{con}$ in EXPERTS = $\text{LOANSTRENGTH}^{dir} - \text{LOANSTRENGTH}^{con}$ in ADVISERS, $\text{LOANSTRENGTH}^{ind} - \text{LOANSTRENGTH}^{con}$ in EXPERTS = $\text{LOANSTRENGTH}^{ind} - \text{LOANSTRENGTH}^{con}$ in ADVISERS, and the same for CUSTOMERS versus ADVISERS. *0.05, **0.01, ***0.001 denote levels of statistical significance.

	EXPERTS			CUSTOMERS		
	CONTROL	DIRECT	INDIRECT	CONTROL	DIRECT	INDIRECT
mean	0.77	−0.75	−0.68	−0.36	−0.93	-0.55
sd	2.52	2.37	2.36	2.84	2.55	2.69
signed-rank z	(3.28)	8.16	8.06	(−2.51)	6.61	3.57
signed-rank p	(0.00)***	0.00***	0.00***	(0.01)**	0.00***	0.00***
diff-in-diff p		0.031*	0.014*		0.579	0.918

analysis of the treatment effects between samples. For this, we compare the treatment effect, for example, of DIRECT in EXPERTS (i.e., $\text{LOANSTRENGTH}^{dir} - \text{LOANSTRENGTH}^{con}$ per participant) with the corresponding treatment effect in sample ADVISERS (i.e., $\text{LOANSTRENGTH}^{dir} - \text{LOANSTRENGTH}^{con}$ per auditor). The bottom row in Table 6 reports the p-values of Wilcoxon rank-sum tests whether the effects of Treatment DIRECT and of Treatment INDIRECT in sample EXPERTS and in sample CUSTOMERS are equal to the corresponding treatment effect in sample ADVISERS in the audit study. Clearly, the experts expected a significantly higher treatment effect of nudges than observed in the audit study (with $p = 0.031$ for DIRECT and $p = 0.014$ for INDIRECT), while the effects expected by customers are statistically not different from the findings of the audit study (with $p = 0.579$ for DIRECT and $p = 0.918$ for INDIRECT).

3.3. Power tests

Recall that we do not find a statistically significant effect in the audit study of the Treatment INDIRECT on LOANPRIO (Result 2). In contrast, the same treatment effect is highly significant (McNemar's $p < 0.001$) in the sample EXPERTS (see Table 5). This raises the question whether the number of observations in the audit study of this particular treatment ($N = 67$) was sufficient to actually detect a treatment effect of the size that was expected by the experts ($N = 122$). To test this, we run paired proportions power tests based on the discordant proportions of LOANPRIO in the sample EXPERTS.

Table 7 reports the results. The confidence with which we can rule out a type II error in the audit study, based on the expected treatment effect of LOANPRIO in the sample EXPERTS, is 94.3% for INDIRECT and 98.9% for DIRECT. This is significantly higher than the traditional power threshold of 80% and gives us confidence that the audit study was sufficiently powered to detect the effect size predicted by the experts. This does not hold for the predictions by potential customers, who expected a much smaller difference between the discordant probabilities of LOANPRIO across the treatments INDIRECT and CONTROL than the experts ($\delta = -0.016$ in sample CUSTOMERS versus $\delta = -0.213$ in sample EXPERTS; see Table 7).

With regard to LOANSTRENGTH, recall that we do not find an effect of nudges in the audit study (Result 3). Power tests show that we should have found treatment effects on LOANSTRENGTH of the size predicted by the experts if they would have existed. As reported in Table 7, the audit study has a power of 93% for treatment DIRECT and 97.4% for treatment INDIRECT, based on

Table 7
Power tests for LOANPRIO and LOANSTRENGTH: The power is computed based on the effect size predictions of the EXPERTS and CUSTOMERS and the observed standard deviations of the corresponding values in the audit study. α is the type I error probability. 'power' is $1-\beta$, where β is the type II error probability. N is the number of pairwise observations in the audit study ($N = 201$, divided by three treatments). δ is the difference between the effects sizes of LOANSTRENGTH and between the discordant probabilities of LOANPRIO. 'v1' and 'v2' are the treatment averages of LOANSTRENGTH and the discordant probabilities of LOANPRIO. 'sddiff' is the standard deviation of the pairwise difference between LOANSTRENGTH^{con} and LOANSTRENGTH^{dir} and between LOANSTRENGTH^{con} and LOANSTRENGTH^{ind}. We use the paired power tests 'power pairedproportions' for LOANPRIO and 'power pairedmeans' for LOANSTRENGTH in Stata. In 'power pairedmeans' we account for known standard deviations of the audit study and apply a finite population correction (fpc). The fpc is $N = 172$, because the number of branches of the two banks in the audit study period was not larger than $N = 343$ (equal to $N = 172$ pairs for paired means). The power for LOANSTRENGTH with an fpc of $N = 343$ is 0.849 for DIRECT V CONTROL and 0.925 for INDIRECT V CONTROL in panel EXPERTS.

		α	power	N	δ	v1	v2	sddiff
EXPERTS								
	DIRECT V CONTROL	0.05	0.989	67	−0.263	0.279	0.016	n/a
LOANPRIO	INDIRECT V CONTROL	0.05	0.943	67	−0.213	0.238	0.025	n/a
	DIRECT V CONTROL	0.05	0.930	67	−0.328	0.770	−0.750	4.639
LOANSTRENGTH	INDIRECT V CONTROL	0.05	0.974	67	−0.373	0.770	−0.680	3.890
CUSTOMERS								
	DIRECT V CONTROL	0.05	0.181	67	−0.065	0.161	0.096	n/a
LOANPRIO	INDIRECT V CONTROL	0.05	0.057	67	−0.016	0.145	0.129	n/a
	DIRECT V CONTROL	0.05	0.251	67	−0.123	−0.360	−0.930	4.639
LOANSTRENGTH	INDIRECT V CONTROL	0.05	0.081	67	−0.049	−0.360	−0.550	3.890

the observed standard deviations in the audit study and on the treatment averages of LOANSTRENGTH as predicted by the sample EXPERTS. For the sample CUSTOMERS, the predicted treatment effects of LOANSTRENGTH are too small to rule out type II errors.

Overall, we can confidently rule out that treatment effects of the magnitude predicted by the experts were missed by chance. We cannot exclude the existence of smaller treatment effects, as predicted by potential customers.

3.4. Process evaluation

It is always possible, particularly with experiments in the field, that unobservable or unintended factors affect the reported results. Moreover, there is a trade-off between internal and external validity, that is, between giving precise unbiased answers to narrow questions and internally less reliable answers to more general questions. For this reason, many clinical trials use 'process evaluations', which has become a gold standard in health-related and medical research (Skivington et al., 2021). Process evaluations typically use qualitative interviews or surveys to provide better insights into contextual factors, which determine and shape whether and how outcomes are generated.

To conduct a process evaluation for our experiment, we administered a hypothetical non pre-registered exploratory survey to 130 financial professionals, whom we recruited via the platform 'behavioral finance online research' (before.world). All respondents work in the financial sector, with an average of 14.8 years of experience, mostly in banks (60%) and in investments (25%). 88% of the respondents report their gender as male (12% female, 0% other). The survey was administered in March and April 2022 (please see Appendix B.5 for the full instructions). In this survey, we evaluated the experiment from several angles.[29]

3.4.1. Long-term relationships

In the experiment, the auditors presented themselves as a one-time customer, who did not have an account at the bank in question. This is different from the usual investor-adviser relationship, which is more long-term. An advisor might be less inclined to promote the loan if long-term reputation plays a role. In our process evaluation survey, we therefore asked financial professionals whether they think that the advice of the bank employee in our Treatment CONTROL would be different, if the auditor would have been a longtime customer of the bank (see Appendix B.5 for details). The majority of the respondents (56.12%) believe that the advice to a long-term customer would not differ much (39.23%) or would be exactly the same (16.92%) for longtime and for first time customers. This indicates that our results may generalize to longtime relationships to a significant extent. Interestingly, the remaining respondents, who believe that the advice would differ for longtime customers (43.85%), are split regarding its direction: 22.31% (21.54%) believe that, for longtime customers, the bank employee would clearly recommend the loan (savings) less. Hence, even when longtime relationships are considered to affect financial advice, it is not automatically in the interest of the customer.

3.4.2. Possible mechanisms

We also attempt to shed more light on four possible mechanisms that may play a role in our treatment effects: moral nudge, reminder, financial literacy, and disciplinary action.

The interventions in both experimental treatments, DIRECT and INDIRECT, can be interpreted as a combination of a moral nudge and a reminder. Moral nudges draw on people's social preferences to follow certain norms or to achieve a positive self-image. They reward "doing the right thing" and thus work through the direct provision of moral (dis)utility.[30] Reminders, in contrast, build on the notion of inattention and limited memory: humans may lose focus or simply forget about their prior intentions during critical moments of decision-making.[31] Both treatments, INDIRECT and DIRECT, combine both components, i.e., the direct utility effect of a moral nudge (adhering to the oath; protecting customer's interests) and the salience effect of a reminder.

We readily acknowledge that we cannot disentangle the two effects in this study. We can, however, split the auditor's nudge into its two components (statement and question), which may give

[29] We thank the reviewer team for proposing this approach.

[30] Moral nudges have been shown to be persistent over time and spill across contexts (Capraro et al., 2019). In Finance, moral nudges have been applied in the field, for example, to enhance tax compliance (Hallsworth et al., 2017) and to increase 401(k) savings rates (Beshears et al., 2015).

[31] In Finance, successful field applications of reminders increased savings rates (Karlan et al., 2016) and loan repayments (Cadena and Schoar, 2011).

us an indication as to their relative importance. For example, in Treatment INDIRECT, it is possible that the auditors' statement (that their current bank cares more about their own profits than about their clients) represents a moral nudge, while the auditors' question (how the adviser's bank protects customers' interests) acts more like a reminder. In an exploratory attempt to disentangle the two effects, we used Treatment INDIRECT from the online survey EXPERTS and split it into two sub-interventions: one that included only the statement, and another one that included only the question (labeled intervention *blue* and *green*, respectively). We administered the two interventions to all respondents within-subject, in a randomized order, and with the same Likert-scales answers as in the audit study (see Appendix B.5 for details). The results show no statistical difference between the two interventions, with a mean of LOANSTRENGTH $= 0.36$ in *blue* and of LOANSTRENGTH $= 0.55$ in *green* (sing-ranked $p = 0.151$ and $z = 1.436$ in a matched-pair Wilcoxon test). Also the outcome frequencies of LOANPRIO in the sub-interventions are similar. A McNemar's test cannot reject the Null that the discordant proportions are equal (McNemar's chi-square $= 0.95$ with $p = 0.33$). Hence, it seems that both components in Treatment INDIRECT, the statement (moral nudge) and the question (reminder), are equally important.

Another mechanism underlying the treatment effect may be that the nudge increases the salience of possible disciplinary action. When a customer mentions the oath, an adviser could infer that such a customer is more likely to file a complaint at the Foundation for Banking Ethics Enforcement. Hence, instead of being morally nudged or reminded to do the right thing, the advisor may be alerted to the potential consequences of breaking the oath.[32] Financial literacy may be another mechanism that is at play. As mentioned in Section 1, previous studies have shown that financially more literate customers receive better financial advice. By mentioning the banker's oath, a client may signal higher financial literacy, which can deter advisers from overly prioritizing bank's interests.

To explore whether the expectation of increased disciplinary action and/or higher perceived financial literacy drive our results, we asked financial professionals which mechanism they think would most likely explain less strong recommendations for loans when customers ask about the purpose of the oath, compared with not asking about the oath (see Appendix B.5 for details). The respondents provide no dominant answer. In total, 47.69% of respondents believe that asking about the oath triggers a sense of moral obligation to do the right thing (25.38%), or that it reminds bank employees to act in the customer's best interests (22.31%). 20% of the respondents think that customers who ask about the oath signal higher financial literacy to whom loans are harder to recommend.[33] 30% of the respondents believe that asking about the oath signals that customers might file a complaint.[34] Hence, none of the four mechanisms (moral nudge, reminder, financial literacy, and disciplinary action) seems to be dominant. In fact, they all play more or less equally important roles (with weights between 20 and 30 percentage points).

4. Conclusion

Since 2015 every employee working in the financial sector in the Netherlands is legally required to take the so-called banker's oath. We study whether nudges that directly or indirectly remind bank employees of their oath affect their financial advice in an ethical dilemma. In a large-scale audit study, we confronted bank employees with a conflict of interest. In the direct nudge treatment, auditors directly addressed advisers about their oath. In the indirect nudge treatment, auditors only indirectly referred to the oath by reminding advisers of customers' interests as its central element. In the control treatment, no nudge was applied. In an additional survey, we elicit the expected results (beliefs) of our audit study from a representative sample of Dutch customers and from Dutch experts in regulation and policy.

We show that, in the control treatment, nearly half of all financial advisers (46.3%) prioritize loans in their recommendations. Direct nudges, however, significantly decrease the likelihood that recommendations prioritize product sales by more than 16 percentage points to only 29.9%. This effect is correctly predicted both by regulation experts and potential customers. Regarding the net strength of advice we find that the direct nudge primarily increases neutral advice without changing the average. Here, experts predict a stronger treatment effect, namely that both direct and indirect nudges significantly shift the average net strength of advice away from loans (product sales). Overall, we find that both experts and customers are correct in assuming that nudges referring to the oath reduce financial advice that prioritizes product sales; experts, however, are wrong in expecting the same effect on the net strength of advice or from nudges that merely remind advisers about customers' interests. Interestingly, customer expectations seem to be closer to our observations in the field than those of experts. On a more speculative note, this might also be an expression of self-serving bias: one could argue that regulatory experts want to believe that the sector, when left to its own devices, will behave immorally, and that the measures they take are effective at changing that behavior.

We find little support for treatment effects of indirect nudges that do not explicitly refer to the banker's oath. This suggests that the banker's oath does play a special role and stands for more than just an increased salience in customers' interests. It is also possible, however, that customer questions about the banker's oath are more surprising than a question about protecting customer interests. Being unexpectedly confronted by questions requiring active engagement may have a stronger effect. In this case it is not the banker's oath itself, but the element of surprise that triggers a response. In line with this, we do find indications that other mechanisms also play a role. Mentioning the oath can, for example, signal higher financial literacy or a higher willingness to file a complaint, both of which increase the likelihood to receive more customer-centered advice. None of these mechanisms, however, seems to be a dominant force in our experimental treatment effects.

The banker's oath is a subject matter that is historically shrouded in wishful thinking (that the oath will jump-start cultural change in the sector), and in ridicule (that the oath is nothing but a ceremonial paper tiger). No single study can provide exhaustive answers, and we hope that our results encourage future research and regulators to take a good look at the banker's oath as a possible policy tool for the financial sector.

Declaration of Competing Interest

None.

[32] Disciplinary action is very rare. In 2016, 2017, 2018, and 2019, the Foundation for Banking Ethics Enforcement ruled in only 3, 11, 24, and 21 cases, respectively (see https://www.tuchtrechtbanken.nl/en/rulings/). Given that the oath is taken by approximately 87,000 bank employees in the Netherlands, an adviser's risk of disciplinary action is small.

[33] We also asked respondents to rate the financial literacy of customers who ask about the oath (Treatment DIRECT) and only about customers' interest but not the oath (Treatment INDIRECT). The scale ranged from 0 (not literate) to 10 (very literate). Customers in DIRECT (INDIRECT) were rated with an average of 5.15 (5.51). The difference is not statistically significant (sign-ranked p = 0.123 and z = -1.540 in a matched-pair Wilcoxon test).

[34] We also asked respondents how likely they consider an average bank customer to file a complaint for breaching the Banker's Oath (between 0% and 100%). The median answer was 10% (i.e., one complaint in ten breaches). 25% (75%) of the respondents considered the likelihood to be smaller or equal to 5% (20%).

Data Availability

Data will be made available on request.

Acknowledgment

We thank Martijn van den Assem, Sascha Füllbrunn, Dennie van Dolder, Aleksander Grocz, Simon Heß, Jürgen Huber, Stefan Palan, Rene Schwaiger, Stephanie Rosenkranz, Janneke Toussaint, Leonard Wolk, Stefan Zeisberger, Wilte Zijlstra, and seminar and conference participants at Aarhus University, University of Innsbruck, Radboud University in Nijmegen, the Dutch Authority for Financial Markets (AFM), the Dutch Ministry of Finance, the Dutch Central Bank, the CPB Netherlands Bureau for Economic Policy Analysis, the Behavioural Insights Netwerk Nederland, and the "Annual Conference of the Bank of Canada: Behavioral Macro-Finance: Implications for Central Bankers", 2021 for very valuable comments. We particularly thank all financial and public institutions involved for their collaboration. Financial support from the Austrian Science Fund FWF (START-grant Y617-G11 and SFB F63), Radboud University, and the Swedish Research Council (grant 2015-01713) is gratefully acknowledged. This study was ethically approved by the IRB of the University of Innsbruck (08/2019) and is pre-registered at https://www.socialscienceregistry.org/trials/4533.

Supplementary material

Supplementary materials of the study including data can be found in the public OSF repository https://osf.io/5s6tu/. ⓡ indicates random order of authors, following Ray and Robson (2018). Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jbankfin.2022.106750.

References

Anagol, S., Cole, S., Sarkar, S., 2017. Understanding the advice of commissions-motivated agents: evidence from the indian life insurance market. Rev Econ Stat 99 (1), 1–15.
Bergstresser, D., Chalmers, J.M.R., Tufano, P., 2009. Assessing the costs and benefits of brokers in the mutual fund industry. Rev Financ Stud 22 (10), 4129–4156.
Bertrand, M., Mullainathan, S., 2004. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. American Economic Review 94 (4), 991–1013.
Beshears, J., Choi, J.J., Laibson, D., Madrian, B.C., Milkman, K.L., 2015. The effect of providing peer information on retirement savings decisions. J Finance 70 (3), 1161–1201.
Boatright, J.R., 2013. Swearing to be virtuous: the prospects of a Banker's oath. Rev Soc Econ 71 (2), 140–165.
Cadena, X., Schoar, A., 2011. Remembering to Pay? Reminders vs. Financial Incentives for Loan Payments. Working Paper. National Bureau of Economic Research.
Capraro, V., Jagfeld, G., Klein, R., Mul, M., de Pol, I.v., 2019. Increasing altruistic and cooperative behaviour with simple moral nudges. Sci Rep 9, 11880.
Carlsson, M., Rooth, D.-O., 2007. Evidence of ethnic discrimination in the swedish labor market using experimental data. Labour Econ 14 (4), 716–729.
Chang, B., Szydlowski, M., 2020. The market for conflicted advice. J Finance 75 (2), 867–903.
Christoffersen, S.E.K., Musto, D.K., 2015. Demand curves and the pricing of money management. Rev Financ Stud 15 (5), 1499–1524.
Cohn, A., Fehr, E., Maréchal, M.A., 2014. Business culture and dishonesty in the banking industry. Nature 516, 86–89.
Darby, M.R., Karni, E., 1973. Free competition and the optimal amount of fraud. The Journal of Law and Economics 16 (1), 67–88.
DellaVigna, S., Pope, D., Vivalt, E., 2019. Predict science to improve science. Science 366 (6464), 428–429.
Dulleck, U., Kerschbamer, R., 2006. On doctors, mechanics, and computer specialists: the economics of credence goods. J Econ Lit 44 (1), 5–42.
Egan, M., 2019. Brokers versus retail investors: conflicting interests and dominated products. J Finance 74 (3), 1217–1260.
Egan, M., Matvos, G., Seru, A., 2019. The market for financial adviser misconduct. J Polit Econ 127 (1), 233–295.
Fecht, F., Hackethal, A., Karabulut, Y., 2018. Is proprietary trading detrimental to retail investors? J Finance 73 (3), 1323–1361.
Fisher, S.R.A., 1935. The Design of Experiments. Macmillan.
Franco, A., Malhotra, N., Simonovits, G., 2014. Publication bias in the social sciences: unlocking the file drawer. Science 345 (6203), 1502–1505.
Gerber, A.S., Green, D.P., 2012. Field Experiments: Design, Analysis, and Interpretation. W.W. Norton, New York.
Guerico, D.D., Reuter, J., 2014. Mutual fund performance and the incentive to generate alpha. J Finance 69 (4), 1673–1704.
Hackethal, A., Inderst, R., Meyer, S., 2012. Trading on advice. Working Paper.
Hallsworth, M., List, J.A., Metcalfe, R.D., Vlaev, I., 2017. The behavioralist as tax collector: using natural field experiments to enhance tax compliance. J Public Econ 148, 14–31.
Harms, J., 2018. Essays on the behavioral economics of social preferences and bounded rationality. Erasmus University Rotterdam.
Hartigan, J.A., Hartigan, P.M., 1985. The dip test of unimodality. Ann. Statist. 13 (1), 70–84.
Heß, S., 2017. Randomization inference with stata: a guide and software. Stata Journal 17 (3).
Hoechle, D., Ruenzi, S., Schaub, N., Schmid, M., 2018. Financial advice and bank profits. Rev Financ Stud 31 (11), 4447–4492.
Imbens, G.W., Rubin, D.B., 2015. Causal Inference in Statistics, Social, and Biomedical Sciences. Cambridge University Press.
Inderst, R., 2010. Misselling (financial) products: the limits for internal compliance. Econ Lett 106 (1), 35–37.
Inderst, R., Ottaviani, M., 2009. Misselling through agents. American Economic Review 99 (3), 883–908.
Inderst, R., Ottaviani, M., 2012. Financial advice. J Econ Lit 50 (2), 494–512.
Inderst, R., Ottaviani, M., 2012. How (not) to pay for advice: a framework for consumer financial protection. J Financ Econ 105 (2), 393–411.
Karlan, D., McConnell, M., Mullainathan, S., Zinman, J., 2016. Getting to the top of mind: how reminders increase saving. Manage Sci 62 (12), 3393–3411.
Loonen, T., Rutgers, M.R., 2017. Swearing to be a good banker: perceptions of the obligatory banker's oath in the netherlands. Journal of Banking Regulation 18, 28–47.
Mullainathan, S., Noeth, M., Schoar, A., 2012. The Market for Financial Advice: An Audit Study. Working Paper. National Bureau of Economic Research.
Oehler, A., Kohlert, D., 2009. Financial advice giving and taking—where are the market's self-healing powers and a functioning legal framework when we need them? Journal of Consumer Policy 32, 91–116.
Ray, D., Robson, A., 2018. Certified random: a new order for coauthorship. American Economic Review 108 (2), 489–520.
Reurink, A., 2018. Financial fraud: a literature review. J Econ Surv 32 (5), 1292–1325.
van Rooij, M., Lusardi, A., Alessie, R., 2011. Financial literacy and stock market participation. J Financ Econ 101 (2), 449–472.
Sharman, J.C., 2010. Shopping for anonymous shell companies: an audit study of anonymity and crime in the international financial system. Journal of Economic Perspectives 24 (4), 127–140.
Skivington, K., Matthews, L., Simpson, S.A., Craig, P., Baird, J., Blazeby, J.M., Boyd, K.A., Craig, N., French, D.P., McIntosh, E., Petticrew, M., Rycroft-Malone, J., White, M., Moore, L., 2021. A new framework for developing and evaluating complex interventions: update of medical research council guidance. BMJ 374.
Stoughton, N.M., Wu, Y., Zechner, J., 2011. Intermediated investment management. J Finance 66 (3), 947–980.