

TI 2021-032/IV
Tinbergen Institute Discussion Paper

The Banker's Oath And Financial Advice

*Utz Weitzel*¹

*Michael Kirchler*²

¹ Vrije Universiteit Amsterdam, Tinbergen Institute

² University of Innsbruck

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and Vrije Universiteit Amsterdam.

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at <https://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900

The Banker’s Oath And Financial Advice

Utz Weitzel*[†]  Michael Kirchler[‡]  [§]

April 9, 2021

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 moral nudges that directly and indirectly 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 moral nudge, referring 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.

JEL: C92, D84, G02, G14

Keywords: experimental finance, audit study, banker’s oath, moral nudges, financial advice.

We thank Sascha Füllbrunn, Aleksander Grocz, Simon Heß, Jürgen Huber, Stefan Palan, Rene Schwaiger, Stephanie Rosenkranz, Janneke Toussaint, Stefan Zeisberger, Wilte Zijlstra, and seminar 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, and the Behavioural Insights Netwerk Nederland 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.socialscisceregistry.org/trials/4533>. Supplementary materials of the study can be found in the public OSF repository <https://osf.io/5s6tu/>.  indicates random order of authors, following Ray and Robson (2018).

*Vrije Universiteit Amsterdam & Tinbergen Institute, De Boelelaan 1105, 1081HV, Amsterdam.

[†]Radboud University, Institute for Management Research, Heyendaalseweg 141, 6525AJ Nijmegen.

[‡]University of Innsbruck, Department of Banking & Finance, Universitätsstrasse 15, 6020 Innsbruck.

[§]Corresponding author: michael.kirchler@uibk.ac.at

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). Moreover, advisers with a history of misconduct can increase their coworkers' propensity to commit financial misconduct, which leads to contagion and a potential exacerbation of misconduct (Dimmock et al., 2018).

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:¹

¹The banker's oath was first introduced on January 1, 2013, but initially restricted to board members and supervisory board members of financial institutions (until 2015). See Loonen and Rutgers (2017) for a detailed account of the introduction of the banker's oath.

“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;
- 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.²

Five years after the general introduction of banker’s oath, this study investigates whether moral 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 groundbreaking 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).³

²For more details on the banker’s oath and its implementation by the independent Foundation for Banking Ethics Enforcement, see <https://www.tuchtrechtbanken.nl/en>.

³Generally, 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,

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.⁴ 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 moral 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. Moral nudges do not rely on bounded rationality or inattention, but 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. 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). 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. In Finance, successful field applications of reminders increased savings rates (Karlan et al., 2016) and loan repayments (Cadena and Schoar, 2011). Our intervention combines both components, i.e., the direct utility effect of a moral nudge with the salience effect of a reminder. We readily acknowledge that we cannot disentangle the two effects in this study. Also, there is the possibility that the intervention reminds advisers of the potential consequences of breaking the oath. An adviser could infer that a customer asking about the oath is more likely to file a complaint at the Foundation for Banking Ethics Enforcement (FBEE). However, disciplinary action is extremely rare. In 2016,

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).

⁴Since 2013, financial advisers and intermediaries in the Netherlands may not be compensated through commissions. This reduces individual incentives for advisers to recommend other products than the requested loan.

2017, 2018, and 2019, the FBEE ruled in 3, 11, 24, and 21 cases, respectively.⁵ Given that the oath is taken by approximately 87,000 bank employees in the Netherlands, an adviser’s risk of disciplinary action is practically negligible.

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 (moral nudge component). Then they asked advisers about how their bank protects customers’ interests (reminder component). After the answer, auditors proceeded as in CONTROL and DIRECT. With this treatment we attempt to disentangle effects that are linked to the oath’s underlying norms and values from its main content. At the same time, this treatment extends to other customer-centered codes of conduct that employers of many banks are asked to sign in acknowledgment without the solemnity of an oath.⁶

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.⁷ 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.

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 moral 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 moral nudges that explicitly refer to the banker’s oath, 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.

⁵See <https://www.tuchtrechtbanken.nl/en/rulings/>

⁶See Boatright (2013) for a detailed discussion of the differences between oaths and professional codes of conduct in banking.

⁷For brevity we henceforth refer to the representative sample of the Dutch working population as ‘customers’, ‘potential customers’ or the sample CUSTOMERS. We readily acknowledge that this sample does not necessarily represent the (potential) customer base of the banks included in our audit study.

When focusing on the expectations of the experts, we find that they predict significant treatment effects for both moral 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). In contrast, the expectations of customers are much closer to the observed effects in the audit study, both with regard to prioritizing loans without intervention (45%) as well as with respect to treatment effects: customers (and experts) both correctly predict that nudges, referring to the banker’s oath, reduce financial advice that prioritizes product sales. However, experts (not customers) are wrong in expecting the same effect from nudges that merely remind advisers about customers’ interests. Overall, the predictions by the experts are less accurate (and more optimistic with regard to the effectiveness of moral 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.

To the best of our knowledge, this is the first study that applies a randomized controlled trial (RCT) to estimate the causal effect of moral 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. The authors also report that advisers encourage returns-chasing behavior and push for investment products with higher agent fees, even if the client starts with a well-diversified, low-fee portfolio. 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, suggesting that it was ineffective in promoting more ethical behavior. We add to this literature by outlining that moral 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 particularly 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 mis-

behavior (Inderst and Ottaviani, 2009; Stoughton et al., 2011; Inderst and Ottaviani, 2012a,b; 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.

The paper is structured as follows. In Section 2 we introduce the experimental setup of the audit study and the prediction tasks, followed by the results in Section 3 and conclusions in Section 4.

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. We sent trained, professional auditors to 201 bank branch offices located all over the Netherlands in 102 cities and villages (see Figure 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.⁸ The auditors impersonated regular customers, who were

⁸In this paper, we do not reveal the names of the two banks, because they are not relevant for the interpretation of the results.

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.⁹

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



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).¹⁰ 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.

⁹All auditors saw the same information material in their preparation for the visits (see the instructions in Appendix B.1).

¹⁰This scenario is a modified version of a conflict of interest used in Harms (2018). Also see Section 2.1.3 for additional details.

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 first applied a moral nudge by asking the bank employees directly about the purpose of the banker’s oath. Specifically, the auditors remarked: “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 as a reminder: “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 as a moral nudge: “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. The difference is subtle but not trivial. A larger treatment effect for DIRECT than for INDIRECT can be interpreted as an indication that the banker’s oath symbolizes and represents more than its (main) content.

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 €3,800 with a minimum of €3,550 (<https://bufferberekenaar.nibud.nl>).¹¹ Hence, there are enough

¹¹NIBUD 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/>

savings to finance the car without the need to take up a loan ($12,000 - 3,800 = 8,200$).¹² 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 advise 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. Figure 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 moral 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 Figure 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 moral nudge in Treatment INDIRECT.

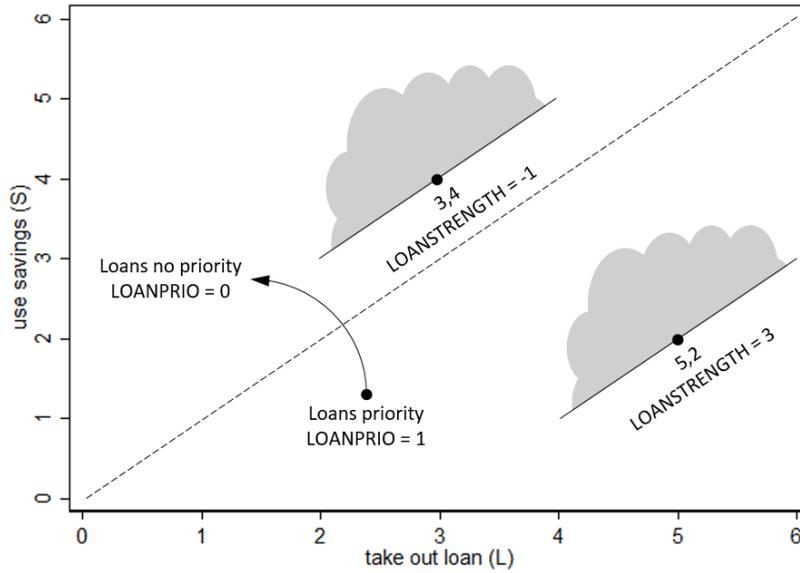
[wp-content/uploads/Nibud-Factsheet-Bereik-en-Imago-2020.pdf](#)). For details on how the buffer is computed see https://www.nibud.nl/wp-content/uploads/A_reference_buffer_for_households.pdf.

¹²Note that the buffer calculation and data collection was completed before the onset of the COVID-19 pandemic where less extreme financial situations were expected.

Hypothesis 1 $\text{LOANPRIO} = 1$ is less likely in Treatment DIRECT than in Treatment CONTROL.

Hypothesis 2 $\text{LOANPRIO} = 1$ is less likely in Treatment INDIRECT than in Treatment CONTROL.

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



As a second outcome measure, we compute the net strength of advice in the bank's interest: $\text{LOANSTRENGTH} = L - S$.¹³ 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. Figure 2 illustrates it with two hypothetical recommendations: one advising to use loans with a net strength of $\text{LOANSTRENGTH} = -1$ and another with $\text{LOANSTRENGTH} = 3$. If a moral 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 Figure 2. For example, if position 5,2 is the advice in favor of loans in the control treatment, then, after a moral nudge, we would expect a shift above the local, solid 45° line (intercepting 5,2) with $\text{LOANSTRENGTH} < 3$. 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 moral nudge increases the strength of advice for loans from 5

¹³Note that $\text{LOANPRIO} = 1$ if $\text{LOANSTRENGTH} > 0$ and $\text{LOANPRIO} = 0$ if $\text{LOANSTRENGTH} \leq 0$.

to 6, but also the advice for using own savings from 2 to 4, then the net advice for using loans has decreased from $\text{LOANSTRENGTH} = 3$ (5,2) to $\text{LOANSTRENGTH} = 2$ (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 moral 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 to the outbreak of COVID-19.¹⁴ 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.

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

¹⁴In 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 meters inter-personal distance in public.

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 balance across treatments and prevent clustering over time.¹⁵ 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.¹⁶ 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.¹⁷ 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 moral 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).

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 (“I knew that already!”), which makes it difficult to appreciate the novelty

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

¹⁶A perfect block randomization per bank was operationally not possible, because of the distances some auditors would have been required to travel.

¹⁷If the adviser was available but did not contact the auditor within 10 minutes, the auditor was allowed to actively approach the adviser.

and contribution of the study (DellaVigna et al., 2019). Despite pre-registration, 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.¹⁸ 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).¹⁹ 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.²⁰ 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.

¹⁸Although 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.

¹⁹Also, for transparency reason and as an additional incentive, they were promised early access to the relevant findings of the audit study.

²⁰For 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.

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 moral nudge of the customers. We are thus confident that the moral 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%).²¹ 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).²² 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.²³ We can therefore conclude that the allocation of advisers to treatments was indeed random with no interference of other factors.

²¹Percentages 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).

²²All 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.

²³Note that all auditor characteristics are balanced across treatments by design (see Section 2.1.3) and are therefore not included in the estimation.

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
SURVEY EXPERTS					
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

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 moral nudges that remind financial advisers of their oath significantly decrease the likelihood that recommendations prioritize product sales (loans).*

Support: Figure 3 reports the fraction of financial advice that prioritizes loans across treatments: the fraction of LOANPRIO = 1 in Treatment CONTROL, where no moral 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.²⁴ In particular, we observe from Figure 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.

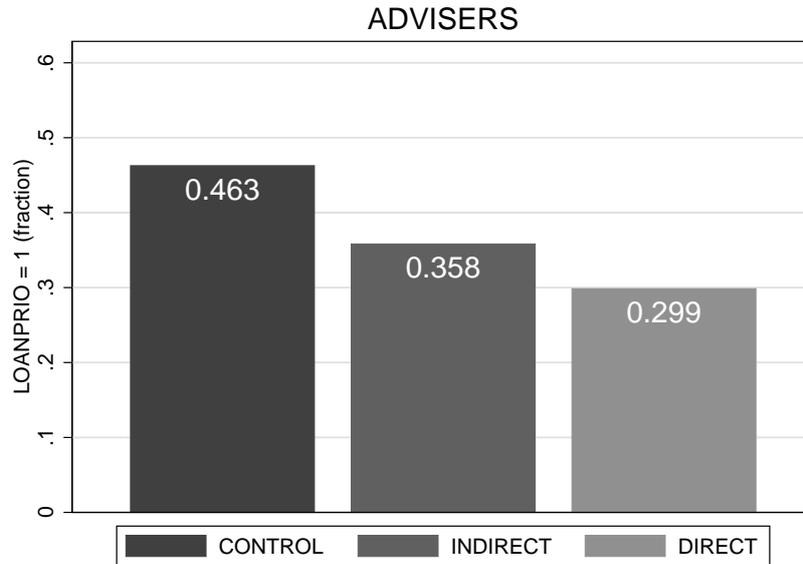
As main analysis, we run a logistic regression with treatment dummies on LOANPRIO as the dependent variable. Models 1 and 2 in Table 2 report the results (odds ratios) with clustered standard errors per auditor. Model 1 regresses the treatment dummies on LOANPRIO. Model 2 accounts for a number of control 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 financial advice is generally less strong.²⁵ In both models, the dummy for the Treatment DIRECT is statistically significant with odds ratios 0.494

²⁴Recall 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.

²⁵All results reported in Table 2 are robust to the exclusion of the variable ‘level’.

and 0.482. This means that, after mentioning the oath ($\text{DIRECT} = 1$), advice that pushes loans ($\text{LOANPRIO} = 1$) is approximately half as likely, in support of Hypothesis 1.

Figure 3: **Fraction of advice to primarily take out a loan, audit study:** CONTROL refers to the control treatment. In Treatment DIRECT, auditors administered a moral 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 ($\text{LOANPRIO} = 1$ if $L > S$).



In addition, we also run permutation tests of all models with 1000 random draws each, accounting for auditor strata.²⁶ 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.²⁷ 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

²⁶We use the user programmed command `ritest` in Stata, described in Heß (2017).

²⁷In 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).

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

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 moral nudge, referring to the oath, significantly decreases the likelihood that recommendations prioritize product sales.

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

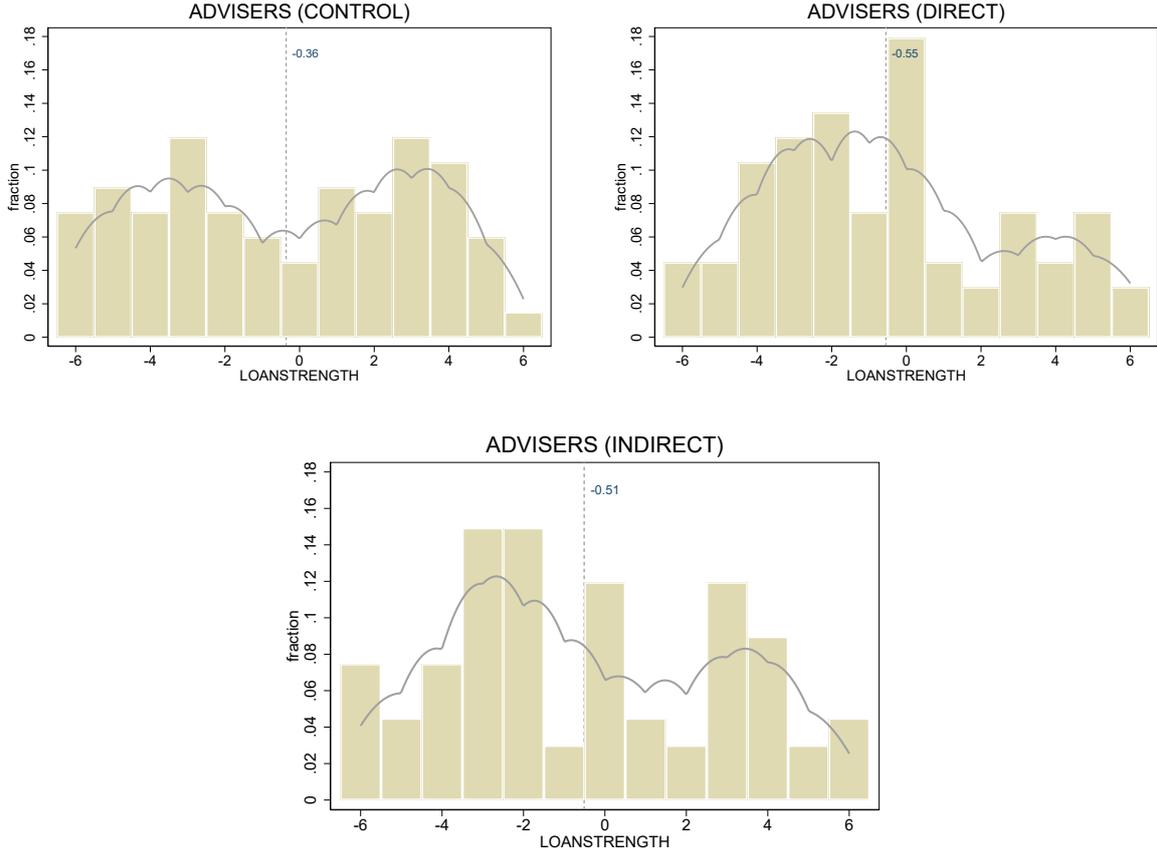
Support: As Figure 3 shows, the unconditional fraction of LOANPRIO = 1 decreases from 0.463 in Treatment CONTROL to 0.358 in Treatment INDIRECT. In the regression analysis, reported in Table 2, the coefficients of the dummy for Treatment INDIRECT in Models 1 and 2 (with LOANPRIO as dependent) are statistically 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. Moral nudges do not decrease net advice to take out loans, but they increase the frequency of neutral advice.*

Figure 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.



Support: Figure 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 $LOANSTRENGTH^{con} = -3$ and $LOANSTRENGTH^{con} = 3$, each with a fraction of 11.9% of all financial advice. As a comparison, neutral advice, which neither favors loans nor savings ($LOANSTRENGTH^{con} = 0$), makes up only 4.5% of all observations.²⁸ As reported in Table A3, $LOANSTRENGTH^{con}$ 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).

²⁸A dip test (Hartigan and Hartigan, 1985) rejects the Null of unimodality with $p = 0.033$.

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.

One reason for this finding is that moral nudges increase neutral financial advice, as shown in Figure 4. In Treatment CONTROL only 4.5% of all advice was neutral ($\text{LOANSTRENGTH}^{\text{con}} = 0$) with two modes at $\text{LOANSTRENGTH}^{\text{con}} = \pm 3$. After the direct moral nudge in Treatment DIRECT, however, $\text{LOANSTRENGTH}^{\text{dir}} = 0$ 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 ($\text{LOANSTRENGTH}^{\text{ind}} = -2$ and $\text{LOANSTRENGTH}^{\text{ind}} = -3$), and, again, $\text{LOANSTRENGTH}^{\text{ind}} = 0$ is more frequent (11.9%) than $\text{LOANSTRENGTH}^{\text{con}} = 0$ (4.5%). To test the effect of moral nudges on neutral advice, we run logistic regressions and permutation tests with a dummy variable for neutral advice ($\text{LOANSTRENGTH} = 0$) as the dependent variable. As reported in Table 4, the direct moral 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. Moral nudges primarily reduce the bimodality of the distribution in Treatment CONTROL, by adding more neutral advice, but have little effect on means and medians.

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 moral nudges significantly reduce the prioritization of product sales. In the field, this effect applies only for direct moral nudges, which is correctly anticipated by customers.*

Support: Figure 5 replicates Figure 3 from the audit study and reports the fraction of $\text{LOANPRIO} = 1$ 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 advisers 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 Figure 3). A two-sided Fisher’s χ^2 exact test 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

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. *0.05 **0.01 ***0.001 denote levels of statistical significance.

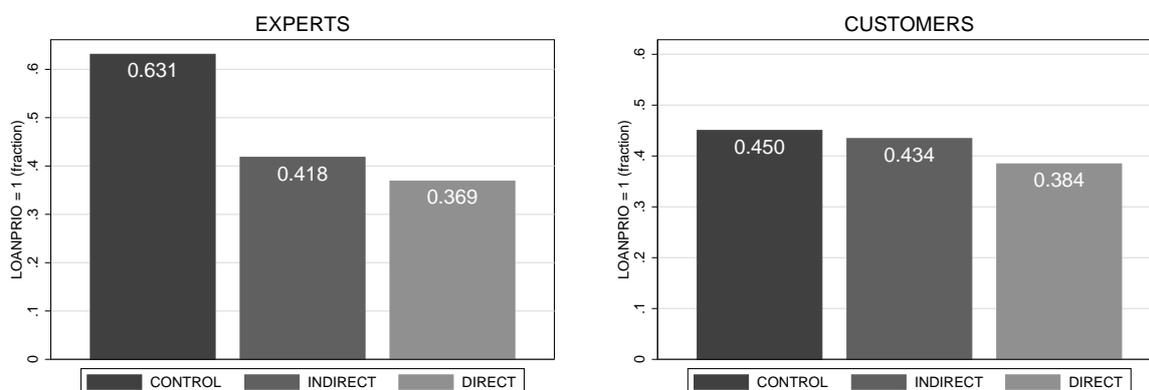
	(1)	(2)
DIRECT	-0.194 (-0.36)	-0.132 (-0.25)
INDIRECT	-0.149 (-0.29)	-0.145 (-0.29)
level		-0.424* (-2.13)
bank		0.509 (1.08)
pos		-0.653* (-2.58)
month		-0.263 (-1.27)
female adv		0.482 (0.89)
age adv		-0.019 (-0.63)
constant	-0.358 (-0.92)	191.979 (1.29)
permute p DIRECT	0.715	0.787
permute p INDIRECT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

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
<i>N</i>	201	201

in CONTROL) is statistically not different from the corresponding fraction in the audit study ($p = 0.474, N = 67 + 502 = 569$).²⁹ 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).

Figure 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.



According to Result 1 from the audit study, direct moral nudges significantly decrease the prioritization of product sales. This finding is correctly anticipated by both experts and potential customers. As shown in Figure 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 $p = 0.496$).

²⁹All other fractions reported in Figure 5 (pertaining to Treatments DIRECT and INDIRECT) are statistically not different from the corresponding fractions in the audit study (see Figure 3).

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
CONTROL	LOANPRIO = 0	43	2	45	42	3	45
	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*	
McNemar	χ^2	28.44			21.13		
	Prob> χ^2	0.000***			0.000***		
CUSTOMERS							
		DIRECT			INDIRECT		
		LOANPRIO = 0	LOANPRIO = 1	Sum	LOANPRIO = 0	LOANPRIO = 1	Sum
CONTROL	LOANPRIO = 0	228	48	276	211	65	276
	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	
McNemar	χ^2	8.44			0.46		
	Prob> χ^2	0.004**			0.496		

Result 5 *Experts predict that moral 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 $p = 0.00$).³⁰ To test whether these predicted effects are actually different from the corresponding effects in the audit study, we run a diff-in-diff analysis of the treatment effects between samples. For this, we compare the treatment effect, for example, of DIRECT in EXPERTS (i.e., $\text{LOANSTRENGTH}^{\text{dir}} - \text{LOANSTRENGTH}^{\text{con}}$ per participant) with the corresponding treatment effect in sample ADVISERS (i.e., $\text{LOANSTRENGTH}^{\text{dir}} - \text{LOANSTRENGTH}^{\text{con}}$ 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 moral nudges than observed in the audit study (with $p = 0.031$ for DIRECT and $p = 0.014$ for INDIRECT), while the effects expected by

³⁰Figure A2 in the Appendix complements Tables A3 and Table 6 by displaying the distributions of LOANSTRENGTH across samples and treatments.

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 $\text{LOANSTRENGTH}^{\text{con}} = \text{LOANSTRENGTH}^{\text{dir}}$ and $\text{LOANSTRENGTH}^{\text{con}} = \text{LOANSTRENGTH}^{\text{ind}}$ within sample. The z - and p -values in parentheses refer to Wilcoxon tests that $\text{LOANSTRENGTH}^{\text{con}}$ 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 $\text{LOANSTRENGTH}^{\text{dir}} - \text{LOANSTRENGTH}^{\text{con}}$ in EXPERTS = $\text{LOANSTRENGTH}^{\text{dir}} - \text{LOANSTRENGTH}^{\text{con}}$ in ADVISERS, $\text{LOANSTRENGTH}^{\text{ind}} - \text{LOANSTRENGTH}^{\text{con}}$ in EXPERTS = $\text{LOANSTRENGTH}^{\text{ind}} - \text{LOANSTRENGTH}^{\text{con}}$ 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

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).

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 $\text{LOANSTRENGTH}^{\text{con}}$ and $\text{LOANSTRENGTH}^{\text{dir}}$ and between $\text{LOANSTRENGTH}^{\text{con}}$ and $\text{LOANSTRENGTH}^{\text{ind}}$. 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								
LOANPRIO	DIRECT v CONTROL	0.05	0.989	67	-0.263	0.279	0.016	n/a
	INDIRECT v CONTROL	0.05	0.943	67	-0.213	0.238	0.025	n/a
LOANSTRENGTH	DIRECT v CONTROL	0.05	0.930	67	-0.328	0.770	-0.750	4.639
	INDIRECT v CONTROL	0.05	0.974	67	-0.373	0.770	-0.680	3.890
CUSTOMERS								
LOANPRIO	DIRECT v CONTROL	0.05	0.181	67	-0.065	0.161	0.096	n/a
	INDIRECT v CONTROL	0.05	0.057	67	-0.016	0.145	0.129	n/a
LOANSTRENGTH	DIRECT v CONTROL	0.05	0.251	67	-0.123	-0.360	-0.930	4.639
	INDIRECT v CONTROL	0.05	0.081	67	-0.049	-0.360	-0.550	3.890

With regard to LOANSTRENGTH, recall that we do not find an effect of moral 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 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.

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 moral 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 implemented a moral nudge that directly reminded advisers of their oath. In the indirect nudge treatment, auditors applied a moral nudge that 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 moral 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 moral 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 (moral) 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.

Also, the fact that a mere reminder of the oath triggers a significant shift toward less product sales (in the interest of the customer) suggests that financial advisers can forget about their pledge, which some have taken many years ago. In discussions with financial advisors and regulators, this aspect was often mentioned as a promising implication of this study: regular reminders of the oath, which has been taken in good faith, are easy to implement and can be an inexpensive method to increase the chances for advice in the customer's interest. Note, however, that the nudges in our study contained an active element on the part of the financial advisor (i.e., being asked to verbalise what the purpose of the oath is, or what the bank does to protect the clients' interests). In an effort to elicit the 'right' behavior from their staff, financial institutions often employ rather passive nudges (stickers, coffee mugs, screen savers, etc.). This raises the

question how effective rather passive nudges really are and how important the active component is in reminding employees of their oath.

Of course, no single study can provide exhaustive answers. We shine some experimental light on a subject matter that is historically shrouded in wishful thinking (that the banker's oath will jump-start cultural change in the sector), and in ridicule (that the banker's oath is nothing but a ceremonial paper tiger). 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.

References

- Anagol, Santosh, Shawn Cole, Shayak Sarkar. 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.
- Bergstresser, Daniel, John M. R. Chalmers, Peter Tufano. 2009. Assessing the costs and benefits of brokers in the mutual fund industry. *The Review of Financial Studies* **22**(10) 4129–4156.
- Bertrand, Marianne, Sendhil Mullainathan. 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, John, James J. Choi, David Laibson, Brigitte C. Madrian, Katherine L. Milkman. 2015. The effect of providing peer information on retirement savings decisions. *The Journal of Finance* **70**(3) 1161–1201.
- Boatright, John R. 2013. Swearing to be virtuous: The prospects of a banker’s oath. *Review of Social Economy* **71**(2) 140–165.
- Cadena, Ximena, Antoinette Schoar. 2011. Remembering to pay? reminders vs. financial incentives for loan payments. Working Paper 17020, National Bureau of Economic Research.
- Capraro, Valerio, Glorianna Jagfeld, Rana Klein, Mathijs Mul, Iris van de Pol. 2019. Increasing altruistic and cooperative behaviour with simple moral nudges. *Scientific Reports* **9** 11880.
- Carlsson, Magnus, Dan-Olof Rooth. 2007. Evidence of ethnic discrimination in the swedish labor market using experimental data. *Labour Economics* **14**(4) 716–729.
- Chang, Briana, Martin Szydlowski. 2020. The market for conflicted advice. *The Journal of Finance* **75**(2) 867–903.
- Christoffersen, Susan E. K., David K. Musto. 2015. Demand curves and the pricing of money management. *The Review of Financial Studies* **15**(5) 1499–1524.
- Cohn, Alain, Ernst Fehr, Michel André Maréchal. 2014. Business culture and dishonesty in the banking industry. *Nature* **516** 86–89.
- Darby, Michael R., Edi Karni. 1973. Free competition and the optimal amount of fraud. *The Journal of Law and Economics* **16**(1) 67–88.
- DellaVigna, Stefano, Devin Pope, Eva Vivalt. 2019. Predict science to improve science. *Science* **366**(6464) 428–429.

- Dimmock, Stephen G., William C. Gerken, Nathaniel P. Graham. 2018. Is fraud contagious? coworker influence on misconduct by financial advisors. *The Journal of Finance* **73**(3) 1417–1450.
- Dulleck, Uwe, Rudolf Kerschbamer. 2006. On doctors, mechanics, and computer specialists: The economics of credence goods. *Journal of Economic Literature* **44**(1) 5–42.
- Egan, Mark. 2019. Brokers versus retail investors: Conflicting interests and dominated products. *The Journal of Finance* **74**(3) 1217–1260.
- Egan, Mark, Gregor Matvos, Amit Seru. 2019. The market for financial adviser misconduct. *Journal of Political Economy* **127**(1) 233–295.
- Fecht, Falko, Andreas Hackethal, Yigitcan Karabulut. 2018. Is proprietary trading detrimental to retail investors? *The Journal of Finance* **73**(3) 1323–1361.
- Fisher, S. R. A. 1935. *The Design of Experiments*. Macmillan.
- Franco, Annie, Neil Malhotra, Gabor Simonovits. 2014. Publication bias in the social sciences: Unlocking the file drawer. *Science* **345**(6203) 1502–1505.
- Gerber, Alan S., Donald P. Green. 2012. *Field Experiments: design, analysis, and interpretation*. W.W. Norton, New York.
- Guerico, Diane Del, Jonathan Reuter. 2014. Mutual fund performance and the incentive to generate alpha. *The Journal of Finance* **69**(4) 1673–1704.
- Hackethal, Andreas, Roman Inderst, Steffen Meyer. 2012. Trading on advice. Working Paper.
- Hallsworth, Michael, John A. List, Robert D. Metcalfe, Ivo Vlaev. 2017. The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics* **148** 14 – 31.
- Harms, Job. 2018. *Essays on the behavioral economics of social preferences and bounded rationality*. ERIM Ph.D. Series Research in Management, Erasmus University Rotterdam.
- Hartigan, J. A., P. M. Hartigan. 1985. The dip test of unimodality. *Ann. Statist.* **13**(1) 70–84.
- Heß, Simon. 2017. Randomization inference with stata: A guide and software. *Stata Journal* **17**(3).
- Hoechle, Daniel, Stefan Ruenzi, Nic Schaub, Markus Schmid. 2018. Financial advice and bank profits. *The Review of Financial Studies* **31**(11) 4447–4492.
- Imbens, G. W., D. B. Rubin. 2015. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

- Inderst, Roman. 2010. Misselling (financial) products: The limits for internal compliance. *Economics Letters* **106**(1) 35 – 37.
- Inderst, Roman, Marco Ottaviani. 2009. Misselling through agents. *American Economic Review* **99**(3) 883–908.
- Inderst, Roman, Marco Ottaviani. 2012a. Financial advice. *Journal of Economic Literature* **50**(2) 494–512.
- Inderst, Roman, Marco Ottaviani. 2012b. How (not) to pay for advice: A framework for consumer financial protection. *Journal of Financial Economics* **105**(2) 393 – 411.
- Karlan, Dean, Margaret McConnell, Sendhil Mullainathan, Jonathan Zinman. 2016. Getting to the top of mind: How reminders increase saving. *Management Science* **62**(12) 3393–3411.
- Loonen, Tom, Mark R Rutgers. 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, Sendhil, Markus Noeth, Antoinette Schoar. 2012. The market for financial advice: An audit study. Working Paper 17929, National Bureau of Economic Research.
- Oehler, Andreas, Daniel Kohlert. 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, Debraj, Arthur Robson. 2018. Certified random: A new order for coauthorship. *American Economic Review* **108**(2) 489–520.
- Reurink, Arjan. 2018. Financial fraud: A literature review. *Journal of Economic Surveys* **32**(5) 1292–1325.
- 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) 127–40.
- Stoughton, Neal M., Youchang Wu, Josef Zechner. 2011. Intermediated investment management. *The Journal of Finance* **66**(3) 947–980.
- van Rooij, Maarten, Annamaria Lusardi, Rob Alessie. 2011. Financial literacy and stock market participation. *Journal of Financial Economics* **101**(2) 449–472.

Online Appendix

A Additional figures and tables

Figure A1: **Financial advice in the control treatment (CONTROL), audit study:** red (blue) circles indicate financial advice with $LOANPRIORITY = 1$ ($LOANPRIORITY = 0$); $N = 67$; white noise added for better display.

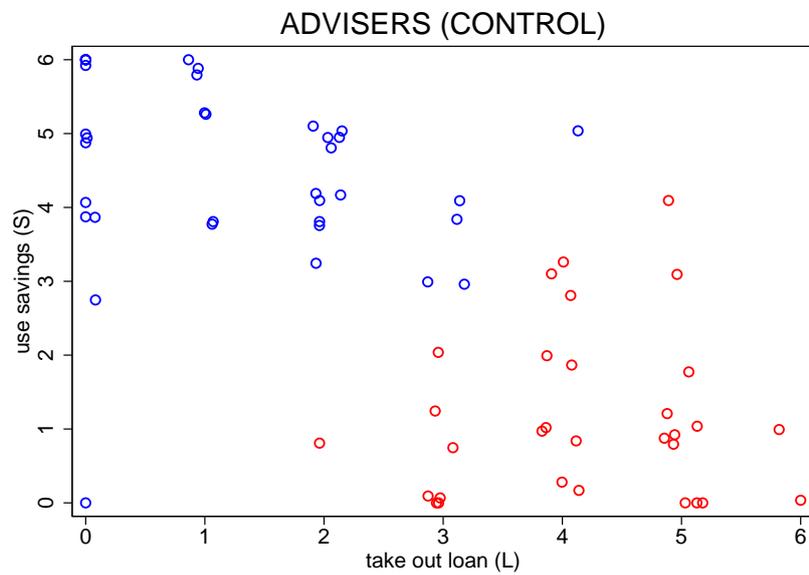


Table A1: **Matched case-control distributions of LOANPRIO, audit study:** 2x2 contingency tables for sample ADVISERS with the outcome frequencies of LOANPRIO in treatments CONTROL v DIRECT and CONTROL v INDIRECT. Observations are matched per auditor-month. ‘Discordant proportions’ refer to switches in LOANPRIO across treatments. (For example, 20 auditors recorded that the advisers prioritized the loan in CONTROL, but the *same* 20 auditors did not receive this advice in DIRECT. The opposite happened to 9 auditors, who experienced LOANPRIO = 0 in CONTROL and LOANPRIO = 1 in DIRECT. The discordant proportions of these switches between the two treatments are $20/67 = 0.299$ and $9/67 = 0.134$, respectively. In $27 + 11 = 38$ cases there was no difference between treatments.) χ^2 and $\text{Prob} > \chi^2$ 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.

ADVISERS							
		DIRECT			INDIRECT		
		LOANPRIO = 0	LOANPRIO = 1	Sum	LOANPRIO = 0	LOANPRIO = 1	Sum
CONTROL	LOANPRIO = 0	27	9	36	28	8	36
	LOANPRIO = 1	20	11	31	15	16	31
Sum		47	20	67	43	24	67
Discordant proportions		0.299	0.134		0.224	0.119	
McNemar	χ^2	4.17			2.13		
	$\text{Prob} > \chi^2$	0.041*			0.144		

Figure A2: Net strength of advice to use own savings (LOANSTRENGTH) across panels: Epanechnikov kernel distributions per treatment. Histogram shows distribution of LOANSTRENGTH in control treatment. Dashed lines show averages of LOANSTRENGTH per treatment.

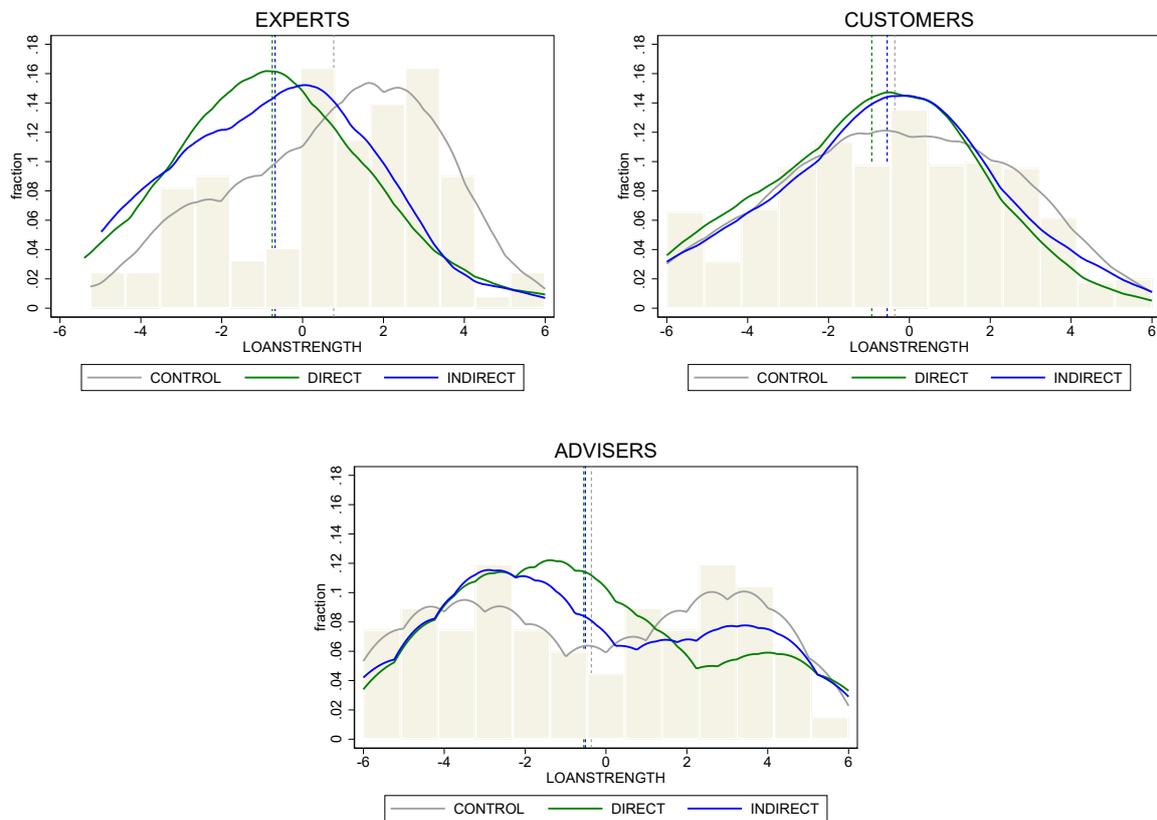


Table A2: **Randomization check, audit study:** multinomial logistic regressions with sample ADVISERS. The dependent is a categorical variable indicating treatments (CONTROL is the baseline; DIRECT and INDIRECT are reported). The month of the branch visit (adviser's age bracket) is included as trend variable (count variable: bracket midpoint) in Model 1 and as a dummy in Model 2. Remaining variables are defined in the notes of Table 1. t-values in parenthesis with standard errors clustered per auditor. *0.05 **0.01 ***0.001 denote levels of statistical significance.

	(1) DIRECT INDIRECT		(2) DIRECT INDIRECT	
bank	0.043 (0.12)	-0.131 (-0.32)	0.078 (0.21)	-0.152 (-0.36)
female adv	-0.152 (-0.44)	0.136 (0.46)	-0.285 (-0.75)	0.027 (0.09)
age adv	-0.012 (-0.71)	-0.009 (-0.48)		
month	0.003 (0.24)	-0.004 (-0.45)		
age adv 31-40			-0.382 (-0.84)	-0.476 (-1.12)
age adv 41-50			0.074 (0.12)	0.030 (0.06)
age adv >51			-0.874 (-1.31)	-0.555 (-0.76)
month 2			0.150 (0.47)	-0.291 (-1.30)
month 3			0.195 (0.63)	-0.242 (-1.42)
month 4			-0.093 (-0.40)	-0.166 (-0.92)
month 5			0.228 (0.71)	-0.317 (-1.17)
month 6			0.008 (0.03)	-0.283 (-1.10)
month 7			0.128 (0.47)	-0.044 (-0.20)
month 8			0.178 (0.92)	-0.003 (-0.02)
constant	-1.788 (-0.18)	3.564 (0.48)	0.266 (0.53)	0.517 (1.22)
Log lik.	-220.098		-218.137	
aic	460.195		484.275	
bic	493		564	
N	201		201	
Clusters	51		51	

Table A3: **Strength of advice (LOANSTRENGTH), audit study**: reports means, standard deviations (sd) and Wilcoxon signed-rank test statistics (signed-rank z, p) for S, L, and LOANSTRENGTH per treatment (CONTROL, DIRECT, INDIRECT) in sample ADVISERS with $N = 67$ observations per cell). S (L) is the strength of advice for using own savings (taking out a loan). LOANSTRENGTH is $L - S$ per respondent/visit. The theoretical range for S and L is from 0 to 6, and from -6 to 6 for LOANSTRENGTH. The z- and p-values (without parentheses) are obtained from Wilcoxon signed-rank tests of the matched pairs $S^{\text{con}} = S^{\text{dir}}$, $S^{\text{con}} = S^{\text{ind}}$, $L^{\text{con}} = L^{\text{dir}}$, $L^{\text{con}} = L^{\text{ind}}$, $\text{LOANSTRENGTH}^{\text{con}} = \text{LOANSTRENGTH}^{\text{dir}}$, and $\text{LOANSTRENGTH}^{\text{con}} = \text{LOANSTRENGTH}^{\text{ind}}$. The z- and p-values in parentheses refer to tests that $\text{LOANSTRENGTH}^{\text{con}}$ is equal to zero (Wilcoxon signed-rank). Observations are matched per auditor-month.

	S			L			LOANSTRENGTH		
	CONTROL	DIRECT	INDIRECT	CONTROL	DIRECT	INDIRECT	CONTROL	DIRECT	INDIRECT
ADVISERS									
mean	2.97	3.15	3.09	2.61	2.6	2.58	-0.36	-0.55	-0.51
sd	2.06	1.85	2.04	1.83	1.71	1.69	3.57	3.23	3.45
signed-rank z		-0.70	-0.71		0.18	0.55	(-0.83)	0.57	0.53
signed-rank p		0.49	0.48		0.86	0.58	(0.41)	0.57	0.59

B Instructions for auditors and prediction task participants

B.1 Instructions for AUDITORS

B.2 Questionnaire for AUDITORS

B.3 Instructions: prediction task for EXPERTS

B.4 Instructions: prediction task for CUSTOMERS

Scenario 1

The **scenario** is the specific question which you will ask the bank employee. This scenario is the same for all visits, but the **treatment** differ from visit to visit. Therefore, carefully read the text below, so that the differences between the scenario and the treatments are clear.

The scenario is as follows:

You want to buy a car of € 8.000,- and you are considering to do this with a loan of the same amount. You are currently employed as an 'uitzend medewerker' and you earn a regular income of € 2.100,- netto per month. You have € 12.000,- in savings, but you consider to keep that in reserve and you do not have any specific plans what to do with that money. You are single, have no children, and you do not have a mortgage or other debts. You are asking the bank about your options.

During each visit you ask the question mentioned above, but you add some extra information for treatment 2 and for treatment 3 (see below).

Treatment 1 (basic treatment)

You use the scenario as describe above. You do not have to add any other information.

- You want to buy a car
- You have € 12.000,- in savings and you consider to keep that money although you do not have specific plans for it
- You want to get a loan of € 8.000,- in order to finance your car
- You want to know your options

Treatment 2 (banker's oath)

You start from the same basic treatment, but you add a **remark about the banker's oath**. You always start the conversation with the remark and question mentioned below, before explaining your scenario:

"I recently saw in a consumer program/ read in the newspaper/ heard from an acquaintance that each bank employee has taken the banker's oath." Then you ask: "What is actually the purpose of the oath?"

- You want to buy a car
- You have € 12.000,- in savings and you consider to keep that money although you do not have specific plans for it
- You want to get a loan of € 8.000,- in order to finance your car
- You want to know your options

Treatment 3 (ethical remark)

You start from the same basic treatment, but you add an **ethical remark with a question**. You always start the conversation with the remark and question mentioned below, before explaining your scenario:

"I have the feeling that my own bank cares more about their own profits than about what is best for their clients." Then you ask: "How does your bank protect the interests of their clients?"

- You want to buy a car
- You have € 12.000,- in savings and you consider to keep that money although you do not have specific plans for it
- You want to get a loan of € 8.000,- in order to finance your car
- You want to know your options

Further take in account the following:

- You have visited the bank branches that have been assigned to you;
- You have fully discussed the case with an advisor and asked the relevant question;
- It is important you do not allow yourself to be 'scared off' with a reference to the website, or a story that they do not offer loans, you clearly state that you would like to be helped on the spot; You can tell the employee that you are willing to move (all) your account(s) to the bank if that is needed to get a loan.
- You have to wait 10 minutes, if an employee has not yet contacted you (and he/she is not in conversation with another customer) then you address the employee yourself. If the employee does not help you immediately, then you wait another 10 minutes. **A conversation must always take place;**
- You prepare yourself in such a way that you actually want to take out a loan and do not indicate that you are merely orientating yourself. In this way you give the consultant the opportunity to manage a purchase.
- Ultimately, it is important that you always get a concrete answer to your question whether a loan is recommended in this situation or whether it is better to use your own savings.

*Important for Treatment 1 and 2: If the employee asks you why you do not get a loan from your own bank, do NOT use the ethical remark as a reason in Treatment 1 and 2. Instead you can say that you want to find the best offer. **The ethical remark should only be used in Treatment 3.***

Questionnaire for auditors

General information

Mystery shopper ID:

Visit ID:

Date:

Exit time:

Entry time:

Treatment applied:

Gender mystery guest:

Male

Female

Age category mystery guest:

21-30 year

31-40 year

41-50 year

51-59 year

Bank:

Were you referred to the advisor by a hostess/host?

Yes/No

SCORE:

Analysis

If treatment 2: Only oath:

Question	1.	Did you clearly ask the oath question as stated in the script before you asked about the financial advice?
Yes		4
No		0

If treatment 2: Only oath:

Question	2.	How much did the employee react to the oath question asked by you?
On a scale from 0 to 6, where 0 is 'Not at all' and 6 is 'Very much'.		

If treatment 3: Only ethical remark:

Question	3.	Did you clearly ask the ethical remark question as stated in the script before you asked about the financial advice?
Yes		4
No		0

If treatment 3: Only ethical remark:

Question	4.	How much did the employee react to the ethical remark question asked by you?
On a scale from 0 to 6, where 0 is 'Not at all' and 6 is 'Very much'.		

Question	5.	To what extent did the employee ask questions to gather further information about your financial situation?
Good		4
Sufficient		3
Neutral		2
Not sufficient		1
Bad		0

Case-specific

Question	6.	How strongly did the advisor recommend to take out a LOAN?						
On a scale from 0 to 6, where 0 is 'Not at all' and 6 is 'Very strongly'.		0 Not at all	1	2	3	4	5	6 Very strongly
		?	?	?	?	?	?	?

Question	7.	How strongly did the advisor recommend to use your SAVINGS?						
On a scale from 0 to 6, where 0 is 'Not at all' and 6 is 'Very strongly'.		0 Not at all	1	2	3	4	5	6 Very strongly
		?	?	?	?	?	?	?

General questions

Question	8.	How busy do you think the advisor was?						
On a scale from 0 to 6, where 0 is 'Not busy at all' and 6 is 'Very busy'.		0 Not at all busy	1	2	3	4	5	6 Very busy
		?	?	?	?	?	?	?

Question	9.	Could other employees or clients overhear your conversation?	
Yes		1	
No		0	

Question	10.	How long (in minutes) was the conversation with the advisor? <i>If no conversation took place, please enter 0 minutes.</i>	
#			

Question	11.	What is the gender of advisor?	
Male		1	
Female		0	

Question	12.	How old do you think the advisor is?	
<20			
21-30			
31-40			
41-50			
>51			

Question	13.	Were there any other things that struck you during the visit?	
Open forum			

Question	14.	Were there any other questions the advisor ask you that you would've liked us to include in the Q&A section?	
Open forum			

Welcome

Welcome

In the last months, we ran a study where professional mystery shoppers visited over 200 branches of Dutch banks.

In this short survey (11 questions) we are interested in your opinion about this mystery shopping study. After the completion of this survey you will get early access to the mystery shopping results.

Participation will take less than 5 minutes.

You will make an important contribution to research. All data will be depersonalized and will only be used for scientific purposes. This study is approved by the ethical commission of Innsbruck University.

Thank you very much for participating!

Utz Weitzel (Professor of Finance: VU Amsterdam, Radboud University)

Michael Kirchler (Professor of Finance: Innsbruck University, Gothenburg University)

*** Please click below to start. Note that you will not be able to go back to previous pages throughout the whole study. ***

These page timer metrics will not be displayed to the recipient.

First Click: *0 seconds*

Last Click: *0 seconds*

Page Submit: *0 seconds*

Click Count: *0 clicks*

General instructions

Background

In the last months, we ran a study where *professional mystery shoppers* visited over 200 branches of Dutch banks. The bank branches were located all over the Netherlands.

The mystery shoppers were disguised as normal customers and approached the bank employees with the following baseline scenario:

- I want to buy a car for € 8.000,- and I am considering to do this with a loan (of the same amount).
- I have € 12.000,- in savings, but I consider to keep that in reserve (and I do not have any specific plans what to do with that money).
- I am currently employed as a temporary worker and I earn a regular income of € 2.100,- net per month.
- I am single, have no children, and I do not have a mortgage or other debts.
- I currently have an account at another bank but I am willing to switch banks.

Then the mystery shoppers asked the bank employee about possibilities and options.

Your task

On the next screen we will ask you for some predictions about the advice of the bank employees.

These page timer metrics will not be displayed to the recipient.

First Click: 0 seconds

Last Click: 0 seconds

Page Submit: 0 seconds

Click Count: 0 clicks

Predictions

This is the main screen with only six questions. Take your time and think carefully about your answers.

On this screen we will ask you to make six predictions about the advice of the bank employees.

- After completion of this survey (in 3 weeks) we will randomly select *one* of your six predictions below and compare it to the *actual* advice as recorded by the mystery shoppers.
- If your selected prediction matches the actual average (+/- 0.2) you earn €25.
- After completion of this survey you will also get early access to the results of our study.

Baseline scenario (reminder)

The mystery shopper:

- wants to buy a car for € 8.000,- and is considering to do this with a loan (of the same amount);
- has € 12.000,- in savings, but considers to keep that in reserve (and does not have any specific plans what to do with that money);
- Is currently employed as a temporary worker and earns a regular income of € 2.100,- net per month;
- is single, has no children, and no mortgage or other debts;
- currently has an account at another bank but is willing to switch banks.

After this scenario, the mystery shopper asked the bank employee about possibilities and options.

What do you predict for this baseline scenario? On average, how strongly did the bank employee recommend to...

	Not at all			Very strongly			
	0	1	2	3	4	5	6
...take out a LOAN?	<input type="text"/>						
...use the SAVINGS?	<input type="text"/>						

Intervention BLUE

In some visits, the mystery shopper added the following to the baseline scenario:

"I recently saw in a consumer program / in the newspaper / heard from an acquaintance that each bank employee has taken the banker's oath."

Then the mystery shopper asked the bank employee:

"What is actually the purpose of the oath?"

After the answer, the mystery shopper asked the bank employee about possibilities and options.

What do you predict for this intervention? On average, how strongly did the bank employee recommend to...

	Not at all			Very strongly			
	0	1	2	3	4	5	6
...take out a LOAN?	<input type="text"/>						
...use the SAVINGS?	<input type="text"/>						

Intervention GREEN

In some visits, the mystery shopper added the following to the baseline scenario:

"I have the feeling that my own bank cares more about their own profits than about what is best for their clients."

Then the mystery shopper asked the bank employee:

"How does your bank protect the interests of their clients?"

After the answer, the mystery shopper asked the bank employee about possibilities and options.

What do you predict for this intervention? On average, how strongly did the bank employee recommend to...

	Not at all					Very strongly
	Not at all	1	2	3	4	Very strongly
	0	1	2	3	4	5
	0	1	2	3	4	5

...take out a LOAN?

...use the SAVINGS?

These page timer metrics will not be displayed to the recipient.

First Click: 0 seconds

Last Click: 0 seconds

Page Submit: 0 seconds

Click Count: 0 clicks

Exit survey

Finally, please answer the following questions:

Which of the following best describes your current job? (several answers possible)

- Regulation
- Policy work
- Analysis
- Research
- Management
- Teaching
- Other (please specify below)

Compared to the average colleague in your organization, how much work experience do you have with projects/topics that are related to our mystery shopping study?

- | | | | | |
|-----------------------|------------------------|-----------------------|------------------------|-----------------------|
| Far below average | Somewhat below average | Average | Somewhat above average | Far above average |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

How much work experience do you have related to regulation and/or policy work in general? (Please enter years of experience)

In which year are you born? (Please enter full calendar year with four digits)

What is your gender?

- Male
- Female
- Other

These page timer metrics will not be displayed to the recipient.

First Click: *0 seconds*

Last Click: *0 seconds*

Page Submit: *0 seconds*

Click Count: *0 clicks*

Block 5

In 3 weeks, you will have early access to the results of the mystery shopping experiment on this web page:

<https://mystshopexp.wordpress.com/>

Please write down the address of the web page so that you can visit it in 3 weeks.

If you want to be eligible for possible payment (in 3 weeks), please enter your email address below (optional).

Entering your email is completely voluntarily and the information will only be used for payment. Your email address will be deleted from the raw data once we completed the payments. After this, all data will be completely de-personalized and cannot be traced back to individuals. All data will be used for academic research purposes only. If you do not enter your email address we assume that you do not want to be eligible for any payment.

Thank you very much for participating.

*** Please click below to complete the survey ***

Welcome

Welcome

Participation will take less than 5 minutes. With your participation you can earn up to €25 (depending on your decisions in this survey). You will receive your payout in points in about 4 weeks.

We are researchers from several universities conducting a survey on your personal opinions and attitudes. You will make an important contribution to research. All data will be depersonalized and will only be used for scientific purposes.

Thank you very much for participating!

Utz Weitzel (Professor of Finance: VU Amsterdam, Radboud University)

Michael Kirchler (Professor of Finance: Innsbruck University, Gothenburg University)

**** Please click below to start. Note that you will not be able to go back to previous pages throughout the whole study. ****

These page timer metrics will not be displayed to the recipient.

First Click: 0 seconds

Last Click: 0 seconds

Page Submit: 0 seconds

Click Count: 0 clicks

General instructions

Background

In the last months, we ran a study where *professional mystery shoppers* visited over 200 branches of Dutch banks. The bank branches were located all over the Netherlands.

The mystery shoppers were disguised as normal customers and approached the bank employees with the following baseline scenario:

- I want to buy a car for € 8.000,- and I am considering to do this with a loan (of the same amount).

- I have € 12.000,- in savings, but I consider to keep that in reserve (and I do not have any specific plans what to do with that money).
- I am currently employed as a temporary worker and I earn a regular income of € 2.100,- net per month.
- I am single, have no children, and I do not have a mortgage or other debts.
- I currently have an account at another bank but I am willing to switch banks.

Then the mystery shoppers asked the bank employee about possibilities and options.

Your task

On the next screen we will ask you for some predictions about the advice of the bank employees.

These page timer metrics will not be displayed to the recipient.

First Click: 0 seconds

Last Click: 0 seconds

Page Submit: 0 seconds

Click Count: 0 clicks

Predictions

This is the main screen with only six questions. Take your time and think carefully about your answers.

On this screen we will ask you to make six predictions about the advice of the bank employees.

- After completion of this survey (in 3 weeks) we will randomly select *one* of your six predictions below and compare it to the *actual* advice as recorded by the mystery shoppers.
- If your selected prediction matches the actual average (+/- 0.2) you earn €25.
- After completion of this survey you will also get early access to the results of our study.

Baseline scenario (reminder)

The mystery shopper:

- *wants to buy a car for € 8.000,- and is considering to do this with a loan (of the same amount);*
- *has € 12.000,- in savings, but considers to keep that in reserve (and does not have any specific plans what to do with that money);*
- *Is currently employed as a temporary worker and earns a regular income of € 2.100,- net per month;*
- *is single, has no children, and no mortgage or other debts;*
- *currently has an account at another bank but is willing to switch banks.*

After this scenario, the mystery shopper asked the bank employee about possibilities and options.

What do you predict for this baseline scenario? On average, how strongly did the bank employee recommend to...

Not at all Very strongly
0 1 2 3 4 5 6

...take out a LOAN?

...use the SAVINGS?

Intervention BLUE

In some visits, the mystery shopper added the following to the baseline scenario:

"I recently saw in a consumer program / in the newspaper / heard from an acquaintance that each bank employee has taken the banker's oath."

Then the mystery shopper asked the bank employee:

"What is actually the purpose of the oath?"

After the answer, the mystery shopper asked the bank employee about possibilities and options.

What do you predict for this intervention? On average, how strongly did the bank employee recommend to...

Not at all Very strongly
0 1 2 3 4 5 6

...take out a LOAN?

...use the SAVINGS?

Intervention GREEN

In some visits, the mystery shopper added the following to the baseline scenario:

"I have the feeling that my own bank cares more about their own profits than about what is best for their clients."

Then the mystery shopper asked the bank employee:

"How does your bank protect the interests of their clients?"

After the answer, the mystery shopper asked the bank employee about possibilities and options.

What do you predict for this intervention? On average, how strongly did the bank employee recommend to...

Not at all Very strongly
0 1 2 3 4 5 6

Not at all 0 1 2 3 4 5 6 Very strongly

...take out a LOAN?

...use the SAVINGS?

These page timer metrics will not be displayed to the recipient.

First Click: 0 seconds

Last Click: 0 seconds

Page Submit: 0 seconds

Click Count: 0 clicks

Exit survey

Finally, please answer the following questions:

Suppose you had €100 in a savings account and the interest rate was 1% per year. After 2 years, how much do you think you would have in the account if you left the money to grow?

- More than € 101
- Less than € 101
- Exactly € 101
- Do not know

Imagine that the inflation was 3% per year and the interest rate on your savings account was 2% per year. After 1 year, how much would you be able to buy with the money in this account?

- Exactly the same
- Less than today
- More than today
- Do not know

Please tell me whether this statement is true or false. "Buying a single company's stock usually provides a safer return than a stock mutual fund."

- False
- True

Do not know

What is your highest level of education?

Have you ever taken out a personal loan at a bank or elsewhere? (for example, a consumer loan or a mortgage)

Yes

No

In which year are you born? (Please enter full calendar year with four digits)

What is your gender?

Male

Female

Other

These page timer metrics will not be displayed to the recipient.

First Click: *0 seconds*

Last Click: *0 seconds*

Page Submit: *0 seconds*

Click Count: *0 clicks*

Block 5

In about 3 weeks, you will receive an email with early access to the results of the mystery shopping experiment.

Thank you very much for participating.

*** Please click below to complete the survey ***