Why do oaths work?

Sorravich Kingsuwankul Chloe Tergiman Marie Claire Villeval

VU Amsterdam & Tinbergen Institute
Smeal College, Penn State University
CNRS, GATE

25 July 2025

PIER Research Workshop
Thailand

Oaths in our society

- Doctors take the Hippocratic Oath.
- Witnesses swear to tell the truth in court.
- Accountants and bankers in some countries swear professional oaths.

Oaths in our society

- Doctors take the Hippocratic Oath.
- Witnesses swear to tell the truth in court.
- Accountants and bankers in some countries swear professional oaths.

Despite the inherent information asymmetry and the lack of legal enforceability of contracts, oaths are considered meaningful.

Oaths in our society

- Doctors take the Hippocratic Oath.
- Witnesses swear to tell the truth in court.
- Accountants and bankers in some countries swear professional oaths.

Despite the inherent information asymmetry and the lack of legal enforceability of contracts, oaths are considered meaningful.

$\rightarrow$ *Why* do oaths work?

A motivating example: the financial markets

- On average, people have low financial literacy and might benefit from better-informed experts.

A motivating example: the financial markets

- On average, people have low financial literacy and might benefit from better-informed experts.
- While trust is important in financial markets, misconducts are not rare:
  - In the United States, 2005-2015: 7% of advisors have misconduct records (15% at some largest advisory firms) (including regulatory offenses, criminal offenses, and customer disputes) (Egan et al. 2019).
  - And, this captures only detected cases...

Open questions

How can we get to more honesty in markets with asymmetric information and where incentives aren't necessarily aligned?

Mullinathan et al. (2012), Anagol et al. (2012), Inderst and Rottaviani (2012), Pool et al. (2016), Foerster et al. (2017), Fecht et al. (2018)

Open questions

How can we get to more honesty in markets with asymmetric information and where incentives aren't necessarily aligned?

Mullinathan et al. (2012), Anagol et al. (2012), Inderst and Rottaviani (2012), Pool et al. (2016), Foerster et al. (2017), Fecht et al. (2018)

- Market mechanisms:
  - Improve detection mechanisms
  - Increase punishment mechanisms
  - Change the incentives

Open questions

How can we get to more honesty in markets with asymmetric information and where incentives aren't necessarily aligned?

Mullinathan et al. (2012), Anagol et al. (2012), Inderst and Rottaviani (2012), Pool et al. (2016), Foerster et al. (2017), Fecht et al. (2018)

- Market mechanisms:
  - Improve detection mechanisms
  - Increase punishment mechanisms
  - Change the incentives

- Non-market mechanisms:
  - Honesty oaths (this paper)

More specifically

- Can oaths be effective? If so, against what kind of lies?
- Do they change advisor decisions? Investor decisions?
- What mechanisms are at play?

More specifically

- Can oaths be effective? If so, against what kind of lies?
- Do they change advisor decisions? Investor decisions?
- What mechanisms are at play?
  - It is intrinsically costly to break a promise: **image concerns**.

More specifically

- Can oaths be effective? If so, against what kind of lies?
- Do they change advisor decisions? Investor decisions?
- What mechanisms are at play?
  - It is intrinsically costly to break a promise: **image concerns**.
    → active audience
    → self as audience
    → passive audience
  - Link to literature on preferences for keeping promises (and for truth-telling).

Literature we relate to: impact of oaths

- Impact of oaths in various laboratory settings: individual decision-making tasks, but also in stategic settings.
- Impact of oaths on practitioners:
  - Weitzel and Kirchler (2023): an audit study showing that reminding Dutch bankers of their oaths makes them less likely to prioritize bank's interests.
  - Heese et al. (2023) exploit the introduction of an integrity pledge for Dutch professional accounting degree holders and document a reduction in discretionary accruals in firms where the CFO took the pledge.

Literature we draw from: preferences for promise keeping and truth telling

- Individuals keep their promises inTrust Games (Ellingsen & Johannesson 2004; Charness & Dufwenberg 2006).
- Vanberg (2008): due to an intrinsic preference for keeping one's words (Ellingsen & Johannesson 2004), not to guilt aversion (Charness & Dufwenberg 2006).
- Individual lying tasks: the decision to lie depends on the intrinsic preference for truth-telling and the reputational cost of being perceived as a liar (Gneezy et al. 2018, Dufwenberg & Dufwenberg 2018, Abeler et al. 2019).
  $\rightarrow$ We pick a setting in which lying is very commonplace (90% lie).

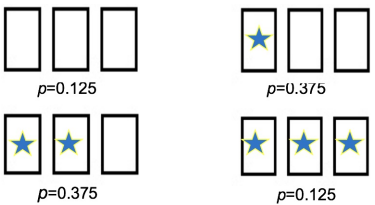
General set-up

General set-up

- Subjects have a (fixed) role of Advisor or Investor.
- Each Advisor is matched with an Investor.
- Information asymmetry: advisors know more than investors.

Advisor

- Advisor receives 3 cards that determine the quality of their portfolio.
- Each card has a 0.5 probability of displaying a star.

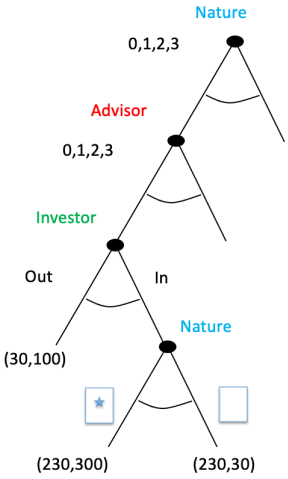


- Face of the cards is private information to the Advisor.
- Advisor's task: send a message m to the Investor announcing their number of stars, $m \in \{0, 1, 2, 3\}$.
- Message m need not match truth τ.

Investor

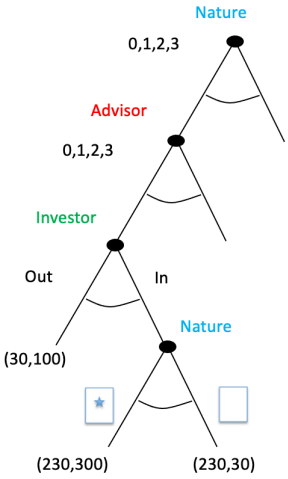
- Investors has 100 tokens.
- Receives Advisor's message m.
- Investor's task: invest or not.
- After the Investor's decision is made:
  - One of the Advisor's three cards is randomly selected.
  - It is shown to both players, regardless of the decision to invest or not.
  - No feedback on the initial portfolio.

Payoffs



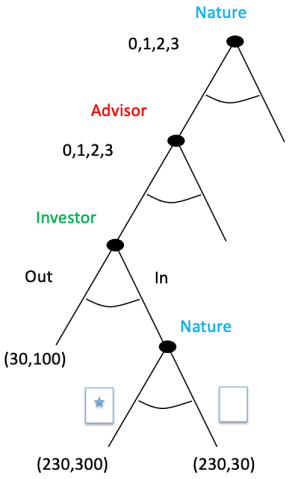
1. Advisor receives 3 cards.

Payoffs



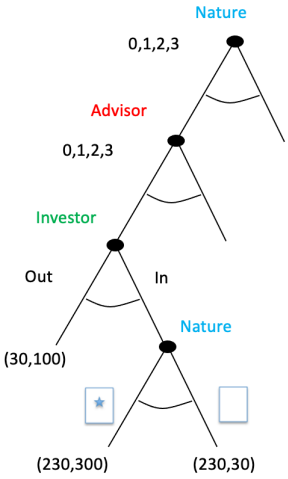
1. Advisor receives 3 cards.
2. Advisor sends message m.

Payoffs



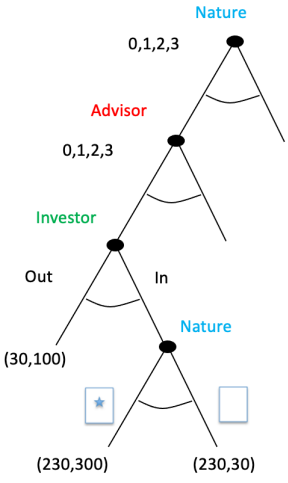
1. Advisor receives 3 cards.
2. Advisor sends message m.
3. Investor invests or not.

Payoffs



1. Advisor receives 3 cards.
2. Advisor sends message m.
3. Investor invests or not.
4. Nature publicly draws one of Advisor's cards.

Payoffs



1. Advisor receives 3 cards.
2. Advisor sends message m.
3. Investor invests or not.
4. Nature publicly draws one of Advisor's cards.

$$\pi^{Advisor} = \begin{cases} 30 & \text{if Investor doesn't invest} \\ 230 & \text{if Investor invests} \end{cases}$$

$$\pi^{I} = \begin{cases} 100 & \text{if doesn't invest} \\ 30 & \text{if invests and Nature draws blank card} \\ 300 & \text{if invests and Nature draws card with star} \end{cases}$$

Theoretical predictions

This is a cheap talk game: a strategy of truth-telling cannot be supported.

In the absence of any information, payoffs are such that in expectation it is financially beneficial for a risk-neutral investor to invest.

Theoretical predictions

This is a cheap talk game: a strategy of truth-telling cannot be supported.

In the absence of any information, payoffs are such that in expectation it is financially beneficial for a risk-neutral investor to invest.

Type of equilibria that emerge are (largely) babbling equilibria:

1. Advisor sends uninformative messages.
2. Investor invests at each period.

Typology of lies – detectable vs. deniable lies

- Detectable lies
  - Lies that can be detected ex-post: announcing 3 stars when actually has fewer.
- Deniable lies
  - Lies that cannot be detected (announcing 2 when the truth is 0 or 1 - cannot for sure know a lie has happened).

Treatments (between-subject design & stranger matching)

- Understanding the impact of oaths on Advisors:
  1. Baseline Treatment: No Oath.
  2. Public-Oath Treatment: Oath is common knowledge:

     "I swear on my honor that during this session I will act honestly and always tell the truth."
  3. Private-Oath Treatment: Oath is private knowledge to the Advisor.
  4. Private-Oath^{++} Treatment: Control treatment; details later.

Treatments (between-subject design & stranger matching)

- Understanding the impact of oaths on Advisors:
  1. Baseline Treatment: No Oath.
  2. Public-Oath Treatment: Oath is common knowledge:

     "I swear on my honor that during this session I will act honestly and always tell the truth."
  3. Private-Oath Treatment: Oath is private knowledge to the Advisor.
  4. Private-Oath^{++} Treatment: Control treatment; details later.
- Understanding the impact of the *knowledge* of oath on Investors:
  5. Additional Treatment 1: details later.
  6. Additional Treatment 2: details later.

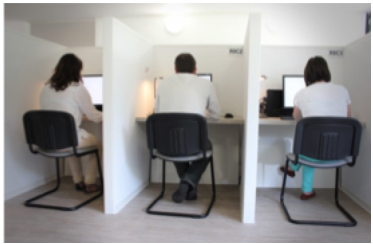
Session details

1. Part 1: Truthful Announcement Game. Everyone has the role of Investor, investment decisions are made with full information in 5 scenarios: 0, 1, 2, 3 and ? stars.
   $\rightarrow$ Reveals subjects' risk preferences.
2. Part 2: Announcement Game. Played 18 times (unknown).
3. Part 3: Belief elicitation.
4. Questionnaire: explanation of their strategy + socio-demographics.

Procedures

- GATE-Lab, Lyon, France.
- 757 subjects (6 treatments). Subjects recruited mainly from local engineering and business schools (HRoot, Bock et al. 2014).
- Average earnings: 20 Euros including a 7-Euro show-up free.

  - zTree (Fischbacher, 2007).
  - 60-75 minutes.



RESULTS - ADVISORS

Impact of a public oath

	No-Oath	Public-Oath	p-value
All Advisors			
% Advisors who Always Tell the Truth	9.2%	74.0%	$p < 0.001$
% Untrue Announcements	52.3%	10.8%	$p < 0.001$

Impact of a public oath

	No-Oath	Public-Oath	p-value
All Advisors			
% Advisors who Always Tell the Truth	9.2%	74.0%	$p < 0.001$
% Untrue Announcements	52.3%	10.8%	$p < 0.001$
Relative Frequency			
Detectable Lies	25.2%	3.9%	$p < 0.001$
Deniable Lies	57.4%	13.8%	$p < 0.001$

Impact of a public oath

	No-Oath	Public-Oath	p-value
All Advisors			
% Advisors who Always Tell the Truth	9.2%	74.0%	$p < 0.001$
% Untrue Announcements	52.3%	10.8%	$p < 0.001$
Relative Frequency			
Detectable Lies	25.2%	3.9%	$p < 0.001$
Deniable Lies	57.4%	13.8%	$p < 0.001$
Among Advisors Who Lie at Least Once			
% Untrue Announcements	57.6%	41.4%	$p = 0.010$
Relative Frequency			
Detectable Lies	27.8%	15.0%	$p = 0.007$
Deniable Lies	63.2%	53.2%	$p = 0.302$

Little detour: investments

Treatments		No-Oath	Public-Oath
0 Stars Announced	% Announcement Game	9.52%	1.69%
	% Truthful Ann. Game	0.0%	2.59%
1 Star Announced	% Announcement Game	14.50%	16.84%
	% Truthful Ann. Game	19.73%	16.88%

Little detour: investments

Treatments		No-Oath	Public-Oath
0 Stars Announced	% Announcement Game	9.52%	1.69%
	% Truthful Ann. Game	0.0%	2.59%
1 Star Announced	% Announcement Game	14.50%	16.84%
	% Truthful Ann. Game	19.73%	16.88%
2 Stars Announced	% Announcement Game	70.51%	85.76%
	% Truthful Ann. Game	98.68%	98.70%
3 Stars Announced	% Announcement Game	77.77%	97.70%
	% Truthful Ann. Game	100.00%	98.70%
Average investment rate		65.8%	56.9%
Average investor earnings		149.7	159.9

Why does the public oath work?

- In addition to monetary payoffs, individuals may bear non-pecuniary costs of breaking a promise:
  - active-audience image costs
  - self-audience image costs
  - passive-audience image costs.
- An oath can impact all these.

Active-audience image costs

- Comparison between Public-Oath and Private-Oath.
- No differences in self- and passive-audience image costs between these treatments.
- Differences indicate whether the common knowledge element is necessary for promise-keeping.

Active-audience image costs

	Public-Oath	Private-Oath	p-value
All advisors			
% Advisors who always tell the truth	74.0%	39.7%	$p < 0.001$
% Untrue announcements	10.8%	27.7%	$p < 0.001$
Relative frequency			
Detectable lies	3.9%	9.6%	$p = 0.002$
Deniable lies	13.8%	37.9%	$p < 0.001$
Among advisors who lie at least once			
% Untrue announcements	41.4%	46.0%	$p = 0.397$
Relative frequency			
Detectable lies	15.0%	15.9%	$p = 0.593$
Deniable lies	53.2%	62.8%	$p = 0.316$

Self- and passive-audience image costs

- Comparison between No-Oath and Private-Oath.
- No differences in active-audience image costs between these treatments.
- In our design, a passive audience can only be the experimenter or god.

Self- and passive-audience image costs

	No-Oath	Private-Oath	p-value
All advisors			
% Advisors who always tell the truth	9.2%	39.7%	$p < 0.001$
% Untrue announcements	52.3%	27.7%	$p < 0.001$
Relative frequency			
Detectable lies	25.2%	9.6%	$p < 0.001$
Deniable lies	57.4%	37.9%	$p = 0.001$
Among advisors who lie at least once			
% Untrue announcements	57.6%	46.0%	$p = 0.004$
Relative frequency			
Detectable lies	27.8%	15.9%	$p = 0.003$
Deniable lies	63.2%	62.8%	$p = 0.884$

Self- or passive-audience?

- Is the change due to concerns about an outside audience or one's own conscience?

Self- or passive-audience?

- Is the change due to concerns about an outside audience or one's own conscience?
- We conducted an additional control treatment (Private-Oath^{++}).
  - Identical to Private-Oath *except* that we removed the identifiability of the decision.
  - We observe negligible differences in Advisors' behaviors across treatments.

Self- or passive-audience?

- Is the change due to concerns about an outside audience or one's own conscience?
- We conducted an additional control treatment (Private-Oath^{++}).
  - Identical to Private-Oath *except* that we removed the identifiability of the decision.
  - We observe negligible differences in Advisors' behaviors across treatments.
  - Corroborates with post-experimental questionnaire, where the modal response (42%) from the truth-tellers referenced their honor or character as it relates to the oath.

Self- or passive-audience?

- Is the change due to concerns about an outside audience or one's own conscience?
- We conducted an additional control treatment (Private-Oath^{++}).
  - Identical to Private-Oath *except* that we removed the identifiability of the decision.
  - We observe negligible differences in Advisors' behaviors across treatments.
  - Corroborates with post-experimental questionnaire, where the modal response (42%) from the truth-tellers referenced their honor or character as it relates to the oath.

This points to the impact of the private oath stemming *primarily* from self-image concerns, with a minor role of a passive audience.

Summary - Advisors

1. Public oath can lead to 75% of advisors to tell the truth (up from 10% when no oath and 40% when oath is private).

2. The mechanism at play concerns three different types of image costs: self, active, and passive audience. Only with common knowledge can we activate all of these.

RESULTS - INVESTORS

Is knowledge of the oath NECESSARY to shift investor behavior?

Additional Treatment 1:

- All subjects are investors.
- Receive an excerpt of the Public-Oath treatment, but without oath part.
- Matched with Advisors from past Public-Oath sessions (so largely honest).

Is knowledge of the oath NECESSARY to shift investor behavior?

Additional Treatment 1:

- All subjects are investors.
- Receive an excerpt of the Public-Oath treatment, but without oath part.
- Matched with Advisors from past Public-Oath sessions (so largely honest).

Announced	Add'l T1	Public-Oath	*p-value*
0 stars	1.1%	2.3%	$p = 0.966$
1 star	11.3%	15.9%	$p = 0.455$
2 stars	84.7%	82.8%	$p = 0.388$
3 stars	95.2%	97.2%	$p = 0.587$
Av. invest rate	56.1%	56.0%	$p = 0.860$

Is knowledge of the oath NECESSARY to shift investor behavior?

Additional Treatment 1:

- All subjects are investors.
- Receive an excerpt of the Public-Oath treatment, but without oath part.
- Matched with Advisors from past Public-Oath sessions (so largely honest).

Announced	Add'l T1	Public-Oath	*p-value*
0 stars	1.1%	2.3%	$p = 0.966$
1 star	11.3%	15.9%	$p = 0.455$
2 stars	84.7%	82.8%	$p = 0.388$
3 stars	95.2%	97.2%	$p = 0.587$
Av. invest rate	56.1%	56.0%	$p = 0.860$

$\rightarrow$ Investors behave just as those in Public-Oath.

$\rightarrow$ Knowledge of the oath is not necessary: experience determines behavior.

Is knowledge of the oath SUFFICIENT to shift investor behavior?

Additional Treatment 2: similar to just described, EXCEPT

- Know about oath, told will be matched with a subset of past advisors.
- Unbeknownst to them, matched with dishonest past advisors ($\sim$ No-Oath).

Is knowledge of the oath SUFFICIENT to shift investor behavior?

Additional Treatment 2: similar to just described, EXCEPT

- Know about oath, told will be matched with a subset of past advisors.
- Unbeknownst to them, matched with dishonest past advisors ($\sim$ No-Oath).

Announced	Add'l T2 (1)	Public-Oath (2)	No-Oath (3)	*p-value* *(1) vs. (2)*	*(1) vs. (3)*
0 stars	0.0%	2.3%	0.0%	$p = 1.000$	$p = 1.000$
1 star	6.6%	15.9%	14.8%	$p = 0.064$	$p = 0.302$
2 stars	64.1%	82.8%	65.9%	$p = 0.002$	$p = 0.947$
3 stars	80.2%	97.2%	76.1%	$p < 0.001$	$p = 0.513$
Av. invest rate	65.1%	56.0%	62.4%	$p = 0.010$	$p = 0.430$

Is knowledge of the oath SUFFICIENT to shift investor behavior?

Additional Treatment 2: similar to just described, EXCEPT

- Know about oath, told will be matched with a subset of past advisors.
- Unbeknownst to them, matched with dishonest past advisors ($\sim$ No-Oath).

Announced	Add'l T2 (1)	Public-Oath (2)	No-Oath (3)	*p-value* *(1) vs. (2)*	*(1) vs. (3)*
0 stars	0.0%	2.3%	0.0%	$p = 1.000$	$p = 1.000$
1 star	6.6%	15.9%	14.8%	$p = 0.064$	$p = 0.302$
2 stars	64.1%	82.8%	65.9%	$p = 0.002$	$p = 0.947$
3 stars	80.2%	97.2%	76.1%	$p < 0.001$	$p = 0.513$
Av. invest rate	65.1%	56.0%	62.4%	$p = 0.010$	$p = 0.430$

$\rightarrow$ By the end, play resembles No-Oath.

$\rightarrow$ Knowledge of oath is not sufficient.

Summary - Investors

1. An honesty oath is not sufficient to induce trust; Investors do not trust blindly.
2. Making the oath common-knowledge is not a necessary condition either for restoring trust ... provided Advisors behave as if Investors were aware of the oath.

Link to field: Prolific survey in Netherlands (N=198)

Banker's oath is required for anyone working in the banking sector.

- Low trust in bankers: 2.55 on a 1-5 scale (doctors: 4, firefighters: 4.6).
- 60% of respondents with high trust say that it was because of good experiences.
- Limited awareness of oath: only 53% are sure or think it is likely that it exists.
- Positive link between awareness and trust (p=0.036).
- 80% of respondents support making these oaths mandatory across Europe, mainly to make clients' interests more salient.

CONCLUSION

CONCLUSION

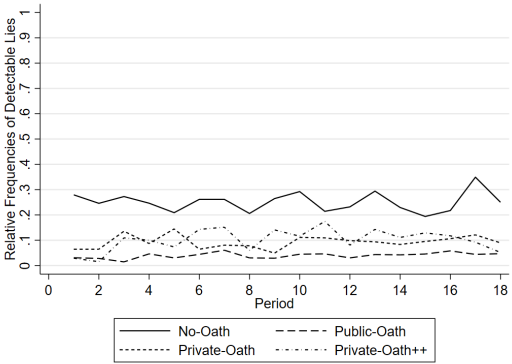
1. Fraud is difficult to deter, even more so if it can be deniable.
2. Non-standard mechanisms like honesty oaths can work against both detectable and deniable lies (even in the absence of sanctions).
3. To have the full impact they need to be common knowledge so both intrinsic private costs as well as reputational costs of breaking the oath are activated.
4. Oaths lead to better investor outcomes as they experience more truthful announcements and then trust them more.

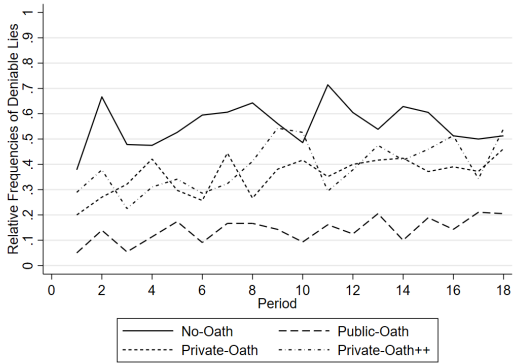
THANK YOU!

s.kingsuwankul@vu.nl

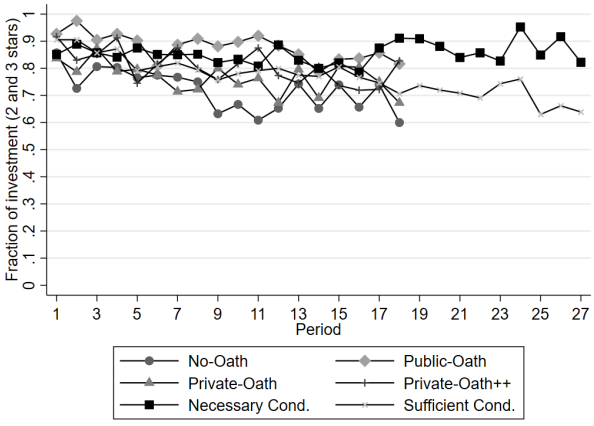
Appendix

Evolution of lies over time





Evolution of investment over time



Evolution of investment over time

Dependent variable Investment decision	(1) No-Oath	(2) Public-Oath	(3) Private-Oath	(4) Private-Oath^{++}	(5) Necessary Cond.	(6) Sufficient Cond.
Period	-0.006*	-0.000	-0.001	-0.002	0.001	-0.007***
	(0.003)	(0.002)	(0.002)	(0.003)	(0.001)	(0.001)
Constants	0.718***	0.574***	0.583***	0.645***	0.540***	0.818***
	(0.032)	(0.027)	(0.034)	(0.034)	(0.028)	(0.027)
Number of observations	1368	1386	1314	1350	2079	2106
Number of clusters	76	77	73	75	77	78

Notes: This table presents the coefficients from random-effects Linear Probability models (GLS) in which the dependent variable is the decision to invest. The only independent variable is the period. Standard errors are clustered at the individual level. * $p<0.05$; ** $p<0.01$; *** $p<0.001$.

Self-audience image concerns drive behavior in Private-Oath

	Private-Oath	Private-Oath++	p-value
All advisors			
% Advisors who always tell the truth	39.7%	38.7%	$p = 0.895$
% Untrue announcements	27.7%	29.2%	$p = 0.734$
Relative frequency			
Detectable lies	9.6%	10.4%	$p = 0.629$
Deniable lies	37.9%	38.1%	$p = 0.896$
Among advisors who lie at least once			
% Untrue announcements	46.0%	47.6%	$p = 0.645$
Relative frequency			
Detectable lies	15.9%	16.9%	$p = 0.593$
Deniable lies	62.8%	62.2%	$p = 0.934$

Literature review

Individual decision-making				
Carlsson et al. (2013)	Contingent valuation	-	No	Decrease of the shares of null and very high willingness-to-pay.
Jacquemet et al. (2013)	Second-price auctions	-	No	Decrease of the shares of null and very high willingness-to-pay.
Krüger (2016)	Coin cheating task	-	Yes	Both positive and negative effects depending on gender
Jacquemet et al. (2017)	Voting referenda	-	No	Decrease of the hypothetical bias.
Heinicke et al. (2019)	Number cheating task	-	Yes	Negative impact on extreme lies, no impact on partial lies, in both gain and loss frames.
Koessler et al. (2019)	Tax evasion game	-	No	No impact on compliance except when combined with non-financial rewards.
Schild et al. (2019)	Mind game	-	Yes	Negative impact when lies are unobservable, no impact when lies are observable.
Beck et al. (2020)	Die cheating task	-	Yes	Negative impact on both extreme and partial lies.
Jacquemet et al. (2020)	Tax evasion game	-	No	No impact on extreme lies, negative impact on partial lies.
Kemper et al. (2020)	Discrete choices	-	Yes	Decrease of hypothetical bias in the estimation of willingness-to-pay.
Prima et al. (2020)	Asset reporting	-	Yes	No impact on lies.
Jacquemet et al. (2021)	Coin cheating task	-	No	Negative impact on extreme lies, no impact on partial lies.
Peer and Feldman (2021)	Reporting perform.	-	Yes	Long-term negative impact on lies.
Akin (2022)	Coin cheating task	-	Yes	Negative effects on undeserved applications.
Babin et al. (2022)	Reporting eye color	-	No	Negative impact on extreme lies, no impact on partial lies.
Babin and Chauhan (2023)	Coin cheating task	-	No	Negative impact on both plausible and implausible misreporting.
Cagala et al. (2024)	Exam-taking	-	Yes	Cheating doubles.
Cagala et al. (2024)	Chip cheating task	-	Yes	No impact of commitment requests on lies in both offline and online settings.
Zickfeld et al. (2025)	Tax evasion game	-	No	Positive impact on tax compliance depending on context.

Literature review

Reference	Task	Common knowledge	Compulsory	Findings
Strategic settings				
Jacquemet et al. (2018)	Coordination game	No	No	Positive impact on coordination rates.
Jacquemet et al. (2019)	Deception game	No	No	For selfish lies in loaded frame, the fraction of liars reduced from 35% in No Oath to 16.7% in (voluntary) Oath. Neutral frame does not reduce lies.
Beck (2021)	Deception game	Yes	Yes	Fraction of liars reduced from 42% in No Oath to 20% in (compulsory) Oath. No impact on the size of lies.
Jacquemet et al. (2021)	Deception game	No	No	For selfish lies, fraction of liars goes from 32.8% in No Oath to 14.3% in (voluntary) Oath. No impact on Pareto lies.
Koessler et al. (2021)	Public good game	Yes	Both	Positive impact of both mandatory and voluntary oaths on contributions, but the effect of the voluntary oath fades away more quickly than when it is mandatory.
Davis and Jaber-Lopez (2022)	Binary social dilemma game	No	Both	Impact of both mandatory and voluntary oaths, but crowding-out effect by non oath takers when it is voluntary.
Hergueux et al. (2022)	Public good game	No	No	Positive impact on contributions, according to social types.
Koessler (2022)	Public good game	Yes	No	Positive impact on contributions. More pledges when the majority of the group does so.
Weitzel and Kirchler (2023)	Audit study	Yes	Yes	Positive impact of reminders of the bankers' oath on real financial advisors' loan recommendations.
Jacquemet et al. (2024)	Trust game	No	No	Oath increases cooperation only with preplay communication. The effect of oath is equivalent to that of a mild fine.

Modeling choices: Active-audience image costs

Two avenues:

- Not wanting to be seen as someone who has broken a promise (judgment on the action) e.g., Gneezy et al. (2018), Abeler et al. (2019), Khalmetski and Sliwka (2019).
- Not wanting to be seen as someone for whom it is easy to break a promise (judgment on the person). e.g., Kartik (2009).

If the first: should see a difference in the reduction of "highly detectable lies" (see 0, say 3) compared with "less detectable lies" (see 2, say 3). However, these are statistically no different.

So, more likely the second.

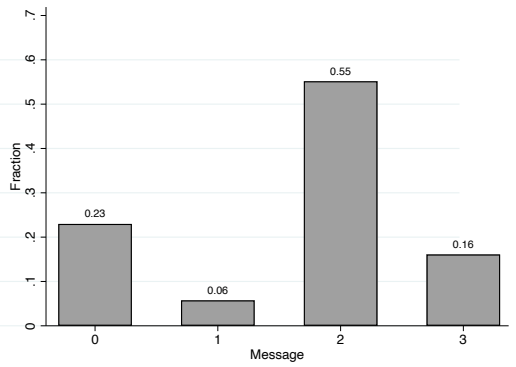
Modeling choices: self-image costs

- Fixed costs?
  - Ruled out.
- Size of the lie: $c = f(|announcement - truth|)$?
  - Our data also allow us to rule this out.
- Instead: seems that people don't want to lie "maximally" if they have promised not to lie at all.

What do self-image costs of breaking an oath look like?

Often modeled as fixed costs or costs that vary with the size of the lie.

- If fixed costs: then oaths shouldn't impact detectable lies more than deniable lies.
  $\rightarrow$ Fraction of $[m = 3|\tau = 0]$ should be greater than $[m = 2|\tau = 0]$.



What do self-image costs of breaking an oath look like?

Often modeled as fixed costs or costs that vary with the size of the lie: $c = f(|m - \tau|)$, f increasing.

- If subject sees 1 star and considers announcing 3:
  $[P_{Inv}(3)x230 + (1 - P_{Inv}(3))x30] \; - \; f(2) \; - \; [(P_{Inv}(1)x230 + (1 - P_{Inv}(1)x30)]$
  $\sim$ *about* $150 - f(2)$.

- If subject sees 0 stars and considers announcing 2:
  $[P_{Inv}(2)x230 + (1 - P_{Inv}(2))x30] \; - \; f(2) \; - \; [(P_{Inv}(0)x230 + (1 - P_{Inv}(0)x30)]$
  $\sim$ *about* $137 - f(2)$.

$\rightarrow$ if anything, we should see more "See 1, say 3" than "See 0, say 2."

$\rightarrow$ Rejected: $P(m = 2|\tau = 0) > P(m = 3|\tau = 1)$
(54% vs. 14%, $p < 0.001$)

$\rightarrow$ Incompatible with size-of-lie costs only.

What do self-image costs of breaking an oath look like?

- Not fixed costs only...
- Not size of the lie only...

Anything constructive?

Interesting pattern in the data:

$P(m{=}3 \mid \text{truth}{=}2) = P(m{=}3 \mid \text{truth}{=}1) = P(m{=}3 \mid \text{truth}{=}0)$

$\rightarrow$ Consistent with fixed cost associated with lying maximally (m=3).