
Supplementary information

Heterogeneity in banker culture and its influence on dishonesty

In the format provided by the
authors and unedited

Zoe Rahwan, Erez Yoeli & Barbara Fasolo

Supplementary Information

Contents

1. Supplementary Discussion
 - 1.1. Differences with the original banker survey
 - 1.2. Differences with the original non-banker survey
2. Supplementary Data
 - 2.1. Bankers**
 - 2.1.1. Descriptive statistics
 - 2.1.2. Randomization checks
 - 2.1.3. Priming checks
 - 2.1.4. Fair coin comparisons
 - 2.1.5. Familiarity checks
 - 2.1.6. Channels for dishonesty
 - 2.2. Non-banking Professionals**
 - 2.2.1. Descriptive statistics
 - 2.2.2. Randomization checks
 - 2.2.3. Priming checks
 - 2.2.4. Treatment effects
 - 2.2.5. Dishonesty estimates
 - 2.2.6. Expectations of dishonesty
 - 2.3. Sample comparisons**
 - 2.3.1. Cross-national cultural analyses
 - 2.3.2. Banker and non-banker analyses
 - 2.3.3. Banker segment analyses
 - 2.4. Experimenter Disclosure**
 - 2.4.1. Descriptive statistics
 - 2.4.2. Randomization checks
 - 2.4.3. Treatment effects
3. Supplementary Notes
 - 3.1. Power Analysis
 - 3.2. Funding
4. Human Subjects Approval
5. Additional References

1.1. Differences with the original banker survey

Due to practical constraints, our studies differed slightly from Cohn et al's. We now describe these differences and postulate their impact.

In introducing the survey, we used incomplete disclosure rather than deception, regarding the purpose of the survey. We draw on the definition of deception that participants "may be given, or be caused to hold, false information about the main purpose of the experiment."³⁹ Participants in our studies were informed that 'norms and attitudes among professionals' were being researched, but not that of honesty explicitly. By contrast, the original study purported to be about 'Life and Satisfaction among Employees.' Such deception was not deemed necessary for the purposes of the research and had compromised possible institutional participation and ethics approval by the Research Ethics Committee of the London School of Economics and Political Sciences. This approach may have made the participants more self-aware and thus less likely to engage in dishonest behaviour. A recent meta-analysis of honesty studies found that the use of deception by experimenters in disclosing the purpose of a study was actually associated with less rather than more dishonesty.⁴⁴ Further, our own experiments reveal no significant differences in dishonesty between participants in incomplete disclosure and deception conditions (SI 1.5).

The potential total value of the reward in both studies was USD 140 - below the USD 200 equivalent used in the original study. The decision to use a lesser amount was guided by consideration of local currency denominations in granting rewards. Despite being less than the original reward, it still provides a substantial financial incentive to over-report the number of winning tosses, especially given that the survey was communicated to take approximately ten minutes. Nevertheless, it is unclear the impact of the lesser financial reward. Traditional economic theory⁵³ would predict lesser cheating. However there is growing evidence that individuals are largely insensitive to rewards.^{40,41} Moreover, some researchers proposing a psychologically-based theory of dishonesty find modest evidence that higher stakes can curtail dishonesty.⁵⁴

The competitive mechanism used to distribute rewards also differed. Cohn et al⁴ made a comparison with a participant drawn from a pilot study. Given the limited time and opportunity to run experiments, the protocol for this research was amended to undertake the comparison with a randomly drawn participant from the same study. Given the reduced psychological distance to the benchmark, this would be expected to exacerbate any effect from the expectations of peer behaviour so it is not problematic. Further, in an unintended error, bankers were only entered into the draw to win their earnings if the total number of winning tosses was greater than a colleagues' rather than greater than or equal to as in the original study. This difference would be expected to result in a greater treatment effect given the competitive nature of banking so again we do not believe this would account for smaller effect sizes.

Unlike the original study, the order of the coin-flipping and investment tasks were not randomised because Cohn et al found no effect from task order in their study. We placed the coin-flipping task ahead of the risk-taking task.

1.2. Differences with the Original Non-banker Survey

Similar to the AP banker sample, for all three non-banker studies we used incomplete disclosure in describing the purpose of the study. Again, participants were informed that 'norms and attitudes among professionals' were being researched. As discussed in SI 1.1 and in presenting our experimental results in SI 1.5, we have diminished concerns that this affected the outcome of our results relative to Cohn et al's studies, all of which deployed deception.

The level of rewards were also the same as the bankers in the Asia Pacific and Middle Eastern jurisdictions: ~USD 14 per coin toss, marking a maximum reward of ~USD 140. This is less than the ~USD 20 per coin toss rewards used in the original study. This is unlikely to drive any difference in reported winning outcomes in light of previous meta-analytical finds of honesty being largely insensitive to rewards,⁴⁰ even at elevated levels for some populations.⁴¹ In terms of field evidence, we find in the AP banker survey that, when charitable rewards were made available - i.e. winnings of up to ~USD 140 would be donated to a charity affiliated with the bank - the level of reported winning outcomes was indistinguishable from when bankers could win rewards for themselves ($p=0.073$, two-sided rank-sum test, Extended Data Figure 7).

In the soft launch of the AP non-banker survey, we identified that at least one banker was in the sample based on their responses to the professional identity priming questions. To ensure the absence of any bankers participating in the survey, for the full launch of the survey we asked in the demographics section of the survey whether participants had any experience in the banking sector, to identify any bankers in the control group. If they answered 'yes', we then asked how many years of experience they had in banking. The full sample for this population ($n=242$) excluded all respondents in both control and treatment groups who identified banking as their current profession ($n=6$) and those in the control group from the soft launch for whom their profession was unknown ($n=11$). These amendments to the survey and subsequent exclusions increased the confidence with which we could describe the sample as representing non-bankers.

In the AP non-banker study, we also sought expectations from the non-bankers regarding the behaviour of other groups. Again, we drew on the questions posed from Cohn et al⁴, asking about expectations for reported winning outcomes in the coin-flipping task for (i) bankers, (ii) prisoners, (iii) doctors and, (iv) the general population and used a between-subject design. In contrast to Cohn et al, we asked these questions to the balanced sample of non-banking professionals, rather than a convenience sample of the general population ($n=183$, males only). We believe this improves representativeness of the findings. These questions were placed after the coin-flipping task, so the participants had the opportunity to learn of the possibility to cheat, and just ahead of the demographic questions, so as to not influence the outcomes of the mock investment task.

2. Supplementary Data

2.1. Banker Data

2.1.1. Descriptive statistics

	Cohn et al		Asia Pacific Bankers		Middle East Bankers	
	Control	Treatment	Control	Treatment	Control	Treatment
Participants (number)	67	61	334	286	77	71
Age (average years)	39.2	38.5	43.6	42.2	35.4	36.5
Males (proportion - %)	59	63	39	41	46	51
Professional Experience (average years)	12.0	10.9	15.3	15.5	10.5	10.3
University Education (proportion - %)	66	57	37	41	90	96
Core Business Unit (proportion - %)	45	52	51	53	55	46

2.1.2. Randomization checks

Bankers were randomly and evenly allocated between treatment and control conditions by the survey platform, Qualtrics. Non-parametric tests were used to assess whether proper randomization had been achieved. Wilcoxon rank sum tests (two-sided) were used for interval variables, while Chi-squared tests (one-sided) were used for binomial variables. The estimates are broadly consistent with randomization being achieved. The only marginal difference was found with regard to age in the treatment and control groups among the Asia Pacific bankers.

a. Asia Pacific Bankers

	Total Sample (n=620 participants)		Control (n=334 participants)		Treatment (n=286 participants)		
Variable	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	p-value
Age (years)^	42.9	0.619	43.6	0.663	42.2	0.453	0.094*
Male	0.40	0.020	0.386	0.027	0.413	0.029	0.504
Professional Experience (years)	15.4	0.457	15.3	0.603	15.5	0.700	0.882

University Education	0.390	0.020	0.374	0.027	0.409	0.029	0.376
Core Business Unit	0.521	0.020	0.515	0.027	0.528	0.030	.747

b. Middle Eastern Bankers

	Total Sample (n=148 participants)		Control (77 participants)		Treatment (71 participants)		
Variable	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	p-value
Age (years)^	36.0	7.13	35.4	6.97	36.5	7.30	0.213
Male	0.71	0.46	0.70	0.46	0.72	0.45	0.821
Professional Experience (years)	10.4	6.31	10.4	6.45	10.3	6.19	0.785
University Education	0.93	0.26	0.90	0.31	0.96	0.20	0.155
Core Business Unit	0.51	0.50	0.55	0.50	0.47	0.50	0.328

* p-value <0.10 ^ For each of the 10-year age brackets, the mid-point was used as a proxy for the age (e.g. if a participant was in the 20-29 age bracket, 25 was used as a proxy).

2.1.3. Priming checks

Bankers were asked to solve six word puzzles, four of which had banking-themed solutions. Comparisons of a number of banking-themed solutions between the control and treatment group is used to assess the effectiveness of priming professional identity.

a. Asia Pacific Bankers

The word puzzles presented to the Asia Pacific bankers were amended from Cohn et al given indications from the Middle Eastern bankers of effects from being a commercial rather than investment bank, and cross-cultural differences. The following banking-themed word puzzles were used:

__v i n g -> saving
c r ____ -> credit
_ o n e y -> money
__ _ n c h -> branch

The number of puzzles solved with banking words was significantly different between the treatment and control group (mean_{treatment} = 1.7, mean_{control} = 1.4, rank-sum test (one-sided), Z=-3.021, p-value=0.001), indicating that the priming of professional identity “worked” and assisted in those treated solving more puzzles with banking-themed solutions.

b. Middle Eastern Bankers

The puzzles solved by the Middle Eastern bankers was drawn directly from Cohn et al. They included the following banking-themed puzzles.

_ _ o c k -> stock
_ _ o k e r -> broker
_ o n e y -> money
B _ n d -> bond

The number of puzzles solved with banking-themed words was not significantly different between the treatment and control group ($\text{mean}_{\text{treatment}} = 1.8$, $\text{mean}_{\text{control}} = 1.7$, rank-sum test (one-sided), $Z = -0.706$, $p\text{-value} = 0.240$), indicating that the priming of professional identity was not effective, unlike in the original study. However, given the similarity in experimental protocol, there is no obvious reason why the priming itself would be ineffective. Rather, it may be that the mechanism used to assess the manipulation was flawed due to unanticipated differences in local vernacular. For example the 'broker' puzzle was commonly solved with 'cooker' - a colloquial word for 'stove' or 'oven' in the region. Further, Cohn et al's study⁴ was conducted at an investment bank, whereas the Middle Eastern institution is a commercial bank. This may explain that only two of 71 subjects in the treatment group solved the '_ _ o c k' word puzzle with 'stock' while the bulk (47) chose 'clock.' Given these considerations, it may be the case that the priming was effective though the measure to test it was flawed.

2.1.4. Fair coin comparisons

As the probability of tossing a winning outcome of a fair coin is known - 0.5 on average - a distribution of winning outcomes from a fair coin over 10 rounds can be calculated. We compare the reported outcomes from our experimental samples with the theoretical distribution of a fair coin to determine if there is any apparent dishonesty. All test results reported below pertain to one-sided rank-sum tests, based on the hypotheses that bankers will cheat more than a fair coin would predict.

a. Asia Pacific Bankers

The treatment group and the control group both report higher winning outcomes than are predicted by a fair coin ($p = 0.002$ and $p = 0.029$, respectively).

b. Middle Eastern Bankers

The treatment group reports higher winning outcomes than are predicted by a fair coin ($p = 0.017$), while the amount of cheating only tended to be significant for the control group ($p = 0.051$).

2.1.5. Familiarity checks

In the AP banker study, we attempted to measure familiarity with the survey tool used. At the end of the survey, we asked the question: Are you familiar with research on banking industry culture which uses a survey like the one you just completed?

Approximately 30% of the respondents answered 'yes.' We found that reporting familiarity with the study could not account for the smaller effect found from priming. Among those that reported being familiar with the study, there is no statistical difference between the winning tosses they reported, whether or not their professional identity was made salient ($\text{mean}_{\text{treated \& familiar}} = 5.6$, $\text{mean}_{\text{control \& familiar}} = 5.5$, $p = 0.160$, rank-sum (one-sided)). Similarly, among the AP bankers reporting being unfamiliar with the study, there is no difference in the reported winning outcomes among the primed bankers and those in the control group ($\text{mean}_{\text{treated \& unfamiliar}} = 5.3$, $\text{mean}_{\text{control \& unfamiliar}} = 5.2$, $p = 0.159$, rank-sum test (one-sided)).

2.1.6. Channels for Dishonesty

Cohn et al⁴ explore various channels through which the salience of professional identity may have increased dishonest behaviour, including materialism, norm obedience and competitiveness (as measured by both self-reports and that associated with varying compensation structures). Given our smaller effect sizes to the extent they become statistically indistinguishable from zero, one would not expect to find evidence for those channels in our studies, and, indeed, we generally did not. We expanded the original study to better understand norm obedience and found some, though not robust, evidence that priming banker identity did change reported expectations of peer behaviour (SI 2.1.6b).

a. Materialism

Materialism, as self-reported on a 1-7 scale, is the only channel for which the authors of the original study find direct empirical support of, noting greater materialism ($p = 0.017$, rank-sum test (one-sided), $n = 128$) among primed bankers and that it is correlated with greater dishonesty (spearman's $\rho_{\text{two-tailed}} = 0.237$, $p = 0.007$). We do not find support for this among our samples, with materialism not increasing significantly with professional identity priming, either among individual samples (AP bankers: $p = 0.575$, ME bankers: $p = 0.118$, rank-sum test (one-sided), $n = 620$ and $n = 148$, respectively) or among AP and ME bankers combined ($p = 0.335$). However, if we extend the sample to include Cohn et al's bankers from the main study, materialism

among primed bankers approaches marginal significance ($\text{mean}_{\text{treatment}} = 4.44$, $\text{mean}_{\text{control}} = 4.29$, $p = 0.111$) and the correlation with reporting winning coin tosses is weak (spearman's $\rho_{\text{two-tailed}} = 0.077$, $p = 0.021$).

b. Norm obedience

Cohn et al⁴ also explore the possibility that bankers' behaviour is affected by beliefs regarding other bankers' behaviour. This belief, in addition to beliefs about what one should do, guides typical norm obedience.⁵⁵ Cohn et al⁴ assess expectations of banker behaviour in the coin-flipping task, by incentivising a separate group of bankers ($n=142$) to predict bank employees' reported winning tosses. In their study, this norm obedience manipulation did not work, and bankers primed with their professional identity did not raise expectations of bankers' dishonesty ($p=0.921$, rank-sum test (two sided)).

In an extension to the original study, we attempted a more direct measure of peer expectations. We asked bankers to report expectations of colleagues' average reported winning tosses in the coin-flipping task, after completing the task. We found some evidence among the ME sample, of primed bankers having marginally greater expectations of their peers reporting greater winning coin tosses ($\text{mean}_{\text{treatment}} = 5.6$, $\text{mean}_{\text{control}} = 5.1$, $p=0.074$, rank-sum test (one-sided), $n=148$) though not in the AP sample ($\text{mean}_{\text{treatment}} = 5.4$, $\text{mean}_{\text{control}} = 5.4$, $p=0.684$, rank-sum test (one-sided), $n=620$). The ME banker finding is potentially indicative of weaker honesty norms associated with banking given the positive correlation with higher winning coin tosses (spearman's $\rho_{\text{two-tailed}} = 0.274$, $p < 0.01$) though may also reflect a post-decision rationalisation of dishonesty or a 'false consensus effect'.⁵⁶

c. Competitiveness

i. Compensation structures

Cohn et al⁴ speculate that priming professional identity may have prompted associations with competitive remuneration schemes - schemes which are more common in core business units as opposed to those providing support services. They find some evidence of elevated cheating among core business units in their probit analysis ($p=0.008$, Wald test), however, they find no interaction with the saliency of professional identity ($p=0.960$, Wald test).

In our regression analyses, across the Asia Pacific and Middle East samples we do not find that working in a 'core unit' is associated with more dishonesty in any of our models (Extended Data Tables 1A and 1B). When pooling banker data from the original and our studies, the core unit is significantly correlated with higher reports of winning coin tosses ($p=0.049$, $p=0.048$, Chi-squared test, in models b and c, respectively), however, similar to Cohn et al, no interaction with priming professional

identity is found ($p=0.702$, $p=0.689$, Chi-squared test, in models b and c, respectively). We note that classifications of bankers into core and support service units are necessarily different in our studies given the different nature of services offered by commercial banks. For consistency, we strived to separate staff conducting support services from other roles (e.g. customer-facing roles).

ii. Self-reports of competitiveness

Cohn et al⁴ also speculate that the culture of banking may be one in which competitiveness is considered to be a desirable behaviour. As such, priming professional identity in combination with the competitive nature of the coin-flipping task, may generate dishonesty among treated bankers. Competitiveness is measured by responses on a seven-point scale to the question: 'How important is it to you to be the best at what you do?' They find no evidence for competitiveness increasing with priming banker identity ($p=0.461$, rank-sum tests (one-sided), $n=128$) and find no correlation between competitiveness (independent of priming) and winning coin flips in their regression analysis ($p=0.642$, Wald test).

Among the AP bankers, higher self-reports of competitiveness were not found among those primed with their professional identity ($p=0.602$, rank sum test (one-sided), $n=620$). In probit analysis, competitiveness was not found to be associated with a higher probability of reporting a winning coin toss (Extended Data Table 1A).

Among the ME bankers, higher self-reports of competitiveness were not found among those primed with their professional identity ($p=0.874$, rank sum test (one-sided), $n=148$). Still, those self-reporting higher levels of competitiveness, independent of priming, were more likely to report a higher number of winning coin flips, according to probit analyses ($p=0.021$, Wald test (Extended Data Table 1B)). No interaction was found between professional identity priming and competitiveness in probit analyses.

Pooling our banker data with Cohn et al, increased competitiveness was not found among bankers primed with their professional identity ($p=0.707$, rank sum test (one-sided), $n=896$), nor was competitiveness (independent of priming) associated with higher reports of winning coin flips (Extended Data Tables 1C and 1D). Together, these results suggest that competitiveness, as measured by self-reports, is not a channel through which priming generated increased dishonesty.

2.2. Non-banking Employees

2.2.1. Descriptive statistics

A balanced sampling technique was used to recruit a population in the Asia Pacific jurisdiction that was representative in terms of gender and age. All participants who joined the study completed screening questions aimed at ensuring that they currently resided in the desired jurisdiction and were employed either part-time or full-time.

a.

	Cohn et al			Asia Pacific		
	Control	Treatment	All	Control	Treatment	All
Participants (number)	66	67	133	125	117	242
Age (average years)	46.0	45.1	45.5	42.2	43.9	43.0
Males (proportion - %)	84.8	91.0	88.0	50.4	48.7	49.6
Professional Experience (average years)	14.1	15.5	14.8	11.5	14.8	13.1
University Education (proportion - %)	84.8	68.7	76.7	47.2	48.7	47.9
Relative Income (7-point scale)	5.52	5.36	5.44	3.90	4.13	4.01

	Middle East			Europe		
	Control	Treatment	All	Control	Treatment	All
Participants (number)	38	29	67	109	96	205
Age (average years)	38.9	42.2	40.4	39.5	40.5	40.0
Males (proportion - %)	60.5	79.3	64.2	46.8	54.2	62.0
Professional Experience (average years)	10.2	9.7	10.0	4.8	6.3	5.5
University Education (proportion - %)	97.4	93.1	95.5	90.8	87.5	89.3
Relative Income (7-point scale)	4.00	4.10	4.05	3.91	4.03	3.97

There are a few apparent differences between the samples of non-bankers. Most noticeably, Cohn et al's⁴ sample had a higher representation of men relative to those

in the Asia Pacific, Middle East and European samples. The original study's sample contained a higher proportion of people with a university education relative to the Asia Pacific sample, though lower than that observed in the Middle East and European samples. The original study's⁴ sample also has more professional experience on average than the other samples, and consistent with that, a perception of higher relative income. None of these variables - gender, university education, professional experience, relative income were found by Cohn et al⁴ to affect the probability of reporting of winning coin tosses.

2.2.2. Randomization checks

Non-banking participants were randomly and evenly allocated between treatment and control conditions by the survey platform, Qualtrics. Non-parametric tests were used to assess whether proper randomisation had been achieved. Wilcoxon rank sum tests (two-sided) were used for interval variables, while Chi-squared tests (one-sided) were used for binomial variables. The estimates are broadly consistent with randomisation being achieved.

a. Asia Pacific Non-bankers

	Total Sample (n=242 participants)		Control (n=125 participants)		Treatment (n=117 participants)		
Variable	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	p-value
Age (average years)	43.0	14.0	42.2	13.7	43.9	14.2	0.399
Males (proportion - %)	49.6	0.501	50.4	0.502	48.7	0.502	0.794
Professional Experience (average years)	13.1	21.6	11.5	11.0	14.8	28.9	0.433
University Education (proportion - %)	47.9	0.501	47.2	0.501	48.7	0.502	0.814
Relative Income (7-point scale)	4.01	1.38	3.90	1.39	4.13	1.36	0.160

b. Middle Eastern Non-bankers

	Total Sample (n=67 participants)		Control (n=38 participants)		Treatment (n=29 participants)		
Variable	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	p-value
Age (average years)	40.4	10.8	38.9	10.0	42.2	11.6	0.193
Males (proportion - %)	68.7	0.467	60.5	0.495	79.3	0.412	0.101
Professional Experience (average years)	10.0	7.5	10.2	6.8	9.7	8.5	0.399
University Education (proportion - %)	95.5	0.208	97.4	0.162	93.1	0.258	0.403
Relative Income (7-point scale)	4.05	1.70	4.00	1.76	4.10	1.66	0.771

c. European Non-bankers

	Total Sample (n=205 participants)		Control (n=109 participants)		Treatment (n=96 participants)		
Variable	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	p-value
Age (average years)	40.0	11.4	39.5	10.6	40.5	12.4	0.600
Males (proportion - %)	50.2	0.501	46.8	0.501	54.2	0.501	0.292
Professional Experience (average years)	5.5	6.1	4.8	5.4	6.3	6.8	0.037*
University Education (proportion - %)	89.3	0.310	90.8	0.290	87.5	0.332	0.443
Relative Income (7-point scale)	3.97	1.32	3.91	1.30	4.03	1.33	0.400

Note that perfect randomisation was not achieved among the European sample of non-bankers with regard to professional experience. Those in the treatment condition self-reported having more years in their industry than those in the control condition. The original study found no relationship between professional experience and the likelihood of reporting a winning outcome in both the main (n=128) and follow-up (n=80) studies.

2.2.3. Priming checks

Similar to the banking professionals, we attempted to assess the effectiveness of professional identity priming among the financial services regulator samples with the use of a word quiz. We did not attempt a manipulation check of the Asia Pacific non-banker sample, in-line with Cohn et al⁴, as we held the concern that the diversity of professions did not lend itself to such an effort.

The use of a quiz as manipulation check for regulators was made more challenging by absence of a previously validated task for regulators and the inability to pre-test the task for effectiveness among regulators - a seemingly more inaccessible population than bankers.

Nevertheless, financial services regulators were asked to solve six word puzzles, four of which had financial and regulatory-themed solutions. Consistent with the bankers, comparisons of the number of professionally-themed solutions between the control and treatment group is used to assess the effectiveness of priming regulatory professional identity.

a. Middle Eastern Financial Services Regulators (Non-bankers)

The word puzzles presented to the Middle Eastern financial services professionals were a combination of two finance-themed puzzles from Cohn et al, and two regulatory-themed puzzles that we developed.⁵⁷

_ o n e y	-> money
B _ n d	-> bond
_ u l e	-> rule
_ i n e	-> fine

The number of puzzles solved was not significantly different between the treatment and control groups ($\text{mean}_{\text{treatment}} = 2.2$, $\text{mean}_{\text{control}} = 2.0$, $p=0.269$, rank-sum test (one-sided)). This may be interpreted to suggest that the induction of professional identity failed, as those in the treatment group did not solve more puzzles. Alternatively, it may indicate that the subjects in the control condition also had a strong financial regulatory identity, independent of priming.

b. European Financial Services Regulators (Non-bankers)

In light of the apparent failure of the manipulation check with the Middle Eastern regulators we reviewed the word puzzle task. Given the low correct response rate to the 'fine' puzzle across both conditions, this puzzle was replaced with the following puzzle: _ a w (law). No other changes were made.

In this sample, we found that participants primed with their professional identity, were slightly more likely to solve word puzzles with finance and regulation-themed words ($\text{mean}_{\text{treatment}} = 1.9$, $\text{mean}_{\text{control}} = 1.7$, $p=0.042$, rank-sum test (one-sided)).

2.2.4. Treatment effects

Non-banker Reported winning outcomes from the coin-flipping task (average)	Control Condition	Treatment Condition	p-value (one-sided rank-sum test)
Asia Pacific (n=242)	5.48	5.85	0.114
Middle East (n=67)	5.11	5.03	0.472
Europe (n=205)	5.26	5.22	0.572

a. Asia Pacific

The mean number of reported wins out of 10 coin flips was 5.85 for the treatment group and 5.48 for the control group, recalling that this sample excluded all participants who self-reported currently working in the banking sector. According to the Wilcoxon rank-sum test (one-sided) [subsequent p-values relate to the same test, unless otherwise stated], the treatment group did not report a statistically different outcome to the control group ($p=0.114$). This result held when removing the 38 participants (16 in the treatment group, 22 in control) who previously (though not currently) held a role in the banking sector ($p=0.293$).

Of note, if we take a subset of participants ($n=38$) who report being bankers with previous experience in banking, a treatment effect is found ($p=0.040$). While this could potentially raise an interesting question regarding the sustainability of internalised norms throughout individuals' careers, our sample is insufficiently powered ($n=16$ in treatment, $n=22$ in control), to make any robust conclusions.

b. Middle East

The mean number of reported winning coin flips from 10 rounds was 5.03 for the treatment group and 5.11 for the control group. The difference between the two groups was not statistically significant ($p=0.472$).

c. Europe

The mean number of reported wins out of 10 coin flips was 5.22 for the treatment group and 5.26 for the control group. The difference in reported outcomes between the two groups was not statistically significantly ($p=0.429$).

2.2.5. Dishonesty estimates

To determine if any statistically detectable dishonesty was occurring in the aggregate level in both conditions, comparisons were made between the theoretical distribution of outcomes from a fair coin and the actual control and treatment group distributions.

a. Asia Pacific.

Participants in both the control and treatment conditions engaged in statistically detectable levels of cheating ($p=0.034$ and $p=0.002$, respectively).

b. Middle East

Relative to the fair coin distribution, neither the control group nor the treatment group reported higher levels of winning outcomes ($p=0.515$ and $p=0.570$, respectively).

c. Europe

Relative to the fair coin, neither the control group nor the treatment group reported higher levels of winning outcomes ($p=0.188$ and $p=0.260$, respectively).

2.2.6. Expectations of dishonesty

We asked the non-banking participants in the Asia Pacific their expectations of reported winning outcomes about different populations, using a between subjects design similar to Cohn et al.⁴ Descriptive statistics are provided in the table below.

a.

Expectations of Reported Outcomes from Coin-flip Task (0-10 wins)		
	Mean	Standard Deviation
Medical Doctors (n=55)	6.04	1.53
Bankers (n=65)	6.34	2.04
Prison Inmates (n=64)	6.23	2.21
General Population (n=58)	5.95	1.78

No differences are found between the expectations of reported outcomes from the coin-flip task ($p\text{-value} = 0.559$, Kruskal-Wallis chi-squared = 2.065) across the four groups. Consistent with this, undertaking pair-wise comparisons between expectations of bankers' reported winnings and those of medical doctors, prison inmates, expectations are not found to differ in any case ($p=0.340$, $p=0.523$,

$p=0.165$, respectively; Bonferroni-corrected significance level $=0.05/3=0.017$). In line with the original study, the pair-wise comparisons are made using two-sided rank-sum tests. This finding contrasts with the original study, which found that expectations banker behaviour in the coin-flipping task were indistinguishable from those of prisoners, tended to be worse than the general population and were significantly worse than for medical doctors.

These differences could also stem from different sampling approaches. We used a panel provider to source a nationally representative sample ($n=242$) based on gender and age, whereas the original study ($n=183$) drew upon a convenience sample of male visitors to a Municipal Office.

We find no strong evidence for heterogeneous gender effects in our sample of Asia Pacific non-bankers. Using two-sided rank-sum tests, we find tentative indications for women holding higher expectations of reported winning outcomes among doctors and prisoners, though these are not significant when correcting for multiple comparisons ($p=0.059$, $p=0.033$; corrected significance level $= 0.05/4 = 0.013$). Further, there is no indication of any gender-based differences for bankers or the general population ($p=0.537$ and $p=0.552$, respectively).

2.3. Sample comparisons

2.3.1. Cross-national cultural analyses

Differences in national culture could be a factor which explains variation in the effect sizes from priming professional identity. Indeed, there is evidence from a study spanning 23 countries that differences in national norms regarding rule violations predict dishonesty.⁵ As such, differences in jurisdictions and their associated national norms, may cause the 'base-line' honesty in each jurisdiction to differ. We make an assessment of that by comparing the non-banker samples and untreated bankers from the undisclosed jurisdiction of the original study, and those in the Asia Pacific and the Middle East, respectively.

In the absence of treatment effects among non-bankers, the treatment and control groups were not separated for the analyses. No differences were found between the samples of non-bankers from the original study ($n=133$), and those from the Asia Pacific ($p=0.542$, two-sided Wilcoxon rank-sum test, $n=242$). In the Middle East, we find that the non-bankers reported less winning outcomes (50.7%) than the non-bankers from the original study (57.8%, $p=0.006$, one-sided Wilcoxon rank-sum test). We are cautious about drawing conclusions from this comparison given the small ($n=67$), non-representative nature of the Middle Eastern study (e.g. only financial services regulatory professionals were surveyed).

Reported Outcomes from Coin-flip Task for Non-bankers (0-10 wins)		
	Mean	Standard Deviation
Cohn et al ($n=133$)	5.78	1.89
Asia Pacific ($n=242$)	5.66	1.91
Middle East ($n=67$)	5.07	1.75

Turning to the untreated bankers in the various jurisdictions, we also find no differences when conducting two-sided rank-sum tests and correcting the p-value for multiple comparisons between those in Cohn's et al's⁴ jurisdiction with those in the Asia Pacific ($p=0.395$) and the Middle East ($p=0.178$) - where dishonesty is detected unlike in the original study.

Further, using measures of rule violations (e.g. national measures of corruption, tax evasion, and fraudulent politics) found to correlate with dishonesty,⁵ we find negligible differences between an advanced Western economy (the presumed original jurisdiction, though is not possible to be confirmed for legal reasons) and our jurisdiction in the Asia Pacific, though notable differences with the Middle East. The latter finding suggests that national cultural differences may have played a role in explaining variation in results between the original jurisdiction and Middle East, as the elevated level of 'baseline' dishonesty among the non-primed bankers may have contributed to a smaller treatment effect.

Overall, these analyses suggest that national differences and honesty norms are unlikely to be responsible for the differences in findings between the AP study and that of the original study, though we cannot so confidently rule out an influence in the Middle Eastern study with the available data.

2.3.2. Banker and Non-banker analyses

We explored whether there were differences in honesty between bankers and non-bankers in each jurisdiction in order to gain an insight into the relative standing of bankers. In comparing the reported winning outcomes of Cohn et al's⁴ unprimed participants, non-bankers reported more winning outcomes than bankers ($p=0.002$). Indeed, Cohn et al⁴ found no evidence that the unprimed bankers engaged in dishonest reporting. By contrast, in the Asia Pacific samples of unprimed bankers and non-bankers, both samples are found to engage in dishonesty when comparing their outcomes to those of a theoretical fair coin distribution ($p=0.029$ and $p=0.021$, respectively). Further, no differences are found in reported winning outcomes for unprimed bankers and non-bankers in the Asia-Pacific (53.0% vs. 54.8%, $p=0.202$).

While less reliable due to the small sample sizes in addition to methodological and sampling variations, in the Middle East we also no find differences between unprimed bankers and non-bankers (54.9% vs. 51.1%, $p=0.109$). We would note that the control group of Middle Eastern bankers tended not report winning outcomes that differed from those of a fair coin ($p=0.051$ - SI 2.1.4), suggesting that people with more honest tendencies may have self-selected into the banking industry in that sample, similar to the original jurisdiction.

We also explored differences between reported winning outcomes among treated bankers in the original study's jurisdiction and treated bankers the Asia Pacific and Middle Eastern jurisdictions. Again, we find no differences ($p=0.206$ and $p=0.765$ respectively, Wilcoxon rank-sum test (two-sided)). Still, we cannot eliminate the possibility and indeed, likelihood of differing national banking norms, nor indeed that of differing commercial and investment banking cultures.

2.3.3. Banker segment analyses

a. Reported Winning Outcomes for Bankers

	Total Sample			Control			Treatment		
Variable (0-10 wins)	Sample size (n)	Mean	S.D.	Sample size (n)	Mean	S.D.	Sam ple size (n)	Mean	S.D.
Cohn et al (main study)	128	5.48	1.81	67	5.16	1.64	61	5.82	1.93
Asia Pacific	620	5.34	1.71	334	5.30	1.67	286	5.40	1.75
Middle East	148	5.59	1.75	77	5.49	1.61	71	5.69	1.89
Commercial Bankers (AP & ME)	768	5.39	1.72	411	5.33	1.66	357	5.45	1.78

To make an assessment of heterogeneity in banking culture across segments of the industry ideally bankers are drawn from the one jurisdiction. While we are unable to do this due to difficulties in recruiting such samples and impossibility to know the original study's jurisdiction, when comparing the available data for investment and commercial bankers reported outcomes in the coin-flipping task, we do not find any statistical differences. Cohn et al's⁴ primed investment bankers do not appear to cheat more than Asia Pacific commercial bankers, Middle Eastern commercial bankers, nor all commercial bankers combined ($p=0.103$, $p=0.383$, $p=0.134$, respectively). When undertaking the same comparisons with the control groups of bankers, again, no differences are found ($p=0.802$, $p=0.911$, $p=0.847$, respectively).

b. Self-Reported Materialism for Bankers

	Total Sample			Control			Treatment		
Variable (1-7 scale)	Sam ple size (n)	Mean	S.D.	Sample size (n)	Mean	S.D.	Sample size (n)	Mean	S.D.
Cohn et al (main study)	128	4.20	1.49	67	3.94	1.48	61	4.49	1.46
Asia Pacific	620	4.30	1.42	334	4.30	1.43	286	4.30	1.40

Middle East	148	4.75	1.78	77	4.57	1.83	71	4.94	1.71
Commercial Bankers (AP & ME)	748	4.39	1.50	411	4.35	1.52	357	4.43	1.49

Cohn et al⁴ speculated “the professional identity prime may have increased dishonesty through an increase in materialistic values.” While we found no evidence for such channels, it may be due to commercial bankers being less materialistic than investment bankers. Accordingly, we compared a self-report measure thought to approximate materialism between the primed bankers from the original study containing investment bankers, and those from the Asia Pacific, Middle East and a combination of the Asia Pacific and Middle East, the latter representing all commercial bankers in our study. We do not detect elevated levels of materialism among Cohn et al’s sample relative to these three primed groups ($p=0.148$, $p=0.955$, $p=0.337$. respectively).

Reference to control groups, also suggests there are differences in this dimension, but not in the expected direction. Unprimed commercial bankers in the Asia Pacific, Middle East and both samples combined self-report higher levels of materialism ($p=0.022$, $p=0.007$, $p=0.013$). Whether the behaviours as opposed to self-reports of commercial bankers reveal greater materialism than investment bankers remains an open question.

c. Self-Reported Competitiveness for Bankers

	Total Sample			Control			Treatment		
Variable (1-7 scale)	Sample size (n)	Mean	S.D.	Sample size (n)	Mean	S.D.	Sample size (n)	Mean	S.D.
Cohn et al (main study)	128	5.61	1.08	67	5.64	1.06	61	5.59	1.12
Asia Pacific	620	6.07	1.05	334	6.10	1.01	286	6.05	1.09
Middle East	148	6.62	0.69	77	6.69	0.63	71	6.54	0.752
Commercial Bankers (AP & ME)	768	6.18	1.01	411	6.21	0.98	357	6.15	1.05

Similar to Cohn et al,⁴ we do not find any evidence that self-reports of competitiveness have a relationship with reporting winning outcomes (see SI 2.1.6.c(ii)). Still, we compare the self-reported competitiveness of Cohn et al's investment bankers to our samples to explore possible differences in banking culture segments. We speculated that, for example, the greater variable pay found among investment bankers, would attract and possibly generate more competitiveness than that found in commercial bankers.

We compared the self-reports commercial bankers in the Asia Pacific, Middle East and aggregate with Cohn et al's investment bankers and found, against expectations, that primed commercial bankers report being more competitive than their investment banking counterparts ($p < 0.01$, $p < 0.01$, $p < 0.01$). The same findings held when making the same comparisons among unprimed bankers ($p < 0.01$, $p < 0.01$, $p < 0.01$). While commercial bankers report being more competitive, their performance in the coin-flipping task with its embedded competitive lottery, did not appear to induce greater competitive behaviour. Consequently, this finding does not assist in explaining the smaller effect size found among commercial bankers.

2.4. Experimenter Disclosure

To understand the effects of varying experimenter disclosure on the outcomes of a coin-flipping task, we experimentally manipulated such disclosures. Refer to Methods for more details.

2.4.1. Descriptive statistics

A total of 925 participants from Amazon Mechanical Turk drawn from the United States of America completed the survey, across the three conditions: deception ($n=315$), incomplete disclosure ($n=309$) and transparency ($n=301$). Of all the participants, 46% were male. The average age was 35.7 years, with a range of 18 to 76 years of age reported. Approximately 60% of participants reported having graduated from university, and nearly 13% of the sample graduating with a higher degree.

2.4.2. Randomization Checks

Participants were randomly and evenly allocated between deception, incomplete disclosure and transparency conditions. Non-parametric tests were used to assess whether proper randomisation had been achieved. Kruskal-Wallis Chi-squared tests were used for binary variables. The estimates are consistent with randomization being achieved.

a.

	Deception (n=315 participants)		Incomplete Disclosure (n=309 participants)		Transparency (n=301 participants)		Kruskal-Wallis Chi-Squared Test
Variable	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	p-value
Age (years)	35.1	11.1	36.7	11.7	35.2	11.6	0.113
Male	0.438	0.497	0.447	0.498	0.489	.501	0.411
MTurk Experience (years)	2.05	1.88	2.02	1.78	1.86	1.66	0.463
University Education	0.635	0.482	0.576	0.495	0.581	0.494	0.253

2.4.3. Treatment Effects

A Kruskal-Wallis Chi-squared test for differences across the conditions found a difference ($p=0.0272$). No ordering effects of cheating increasing with the transparency of experimenter disclosure were found ($p=0.194$, Jonckheere-Terpstra test (two-sided)).

Comparisons were made between each of the experimental groups, using a Bonferroni-corrected significance level (i.e. $0.05/3 = 0.017$) for Wilcoxon rank-sum tests (one-sided). No differences were found between the deception and incomplete disclosure conditions ($p=0.188$) - the experimental disclosures that Cohn et al⁴ and we used, respectively.

The only difference found was between the incomplete disclosure and transparency conditions ($p=0.006$). Still the size of the difference between the average means that there would have been practically no impact on actually payoffs in the experiment. That is, the difference would translate to a negligible change in financial payoffs after accounting for rounding of reported outcomes (*see table below*).

	Deception (n=315 participants)		Incomplete Disclosure (n=309 participants)		Transparency (n=301 participants)	
Variable	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Reported Winning Outcomes (from 10 coin flips)	5.84	2.13	5.99	1.95	5.64	2.10

Of course there are limitations around generalizing this finding. This study was conducted on a sample of US Amazon Mechanical Turk workers and therefore, may not hold among bankers or other populations. It could be the case that such Amazon Mechanical Turk workers, given their experience participating in many studies and/or pressure to complete the study in a timely fashion⁵⁸, are less attentive to the disclosure purpose of an experiment. Consequently, this is an avenue for future research.

3. Supplementary Notes

3.1. Power Analysis

a. Bootstrapped samples - statistical significance

We determined the likelihood of finding a statistically significant result from rank-sum tests from simulated samples. Samples were drawn with replacement in 50 subject increments up to a sample size of 1000. The samples were drawn from Cohn et al's main study⁴, where a treatment effect was detected.

An even number of bankers were drawn from the control and treatment groups, in keeping with the random allocation of the experimental design. For each sample size, 10,000 simulated samples were drawn and rank-sum tests conducted. The proportion of statistically significant outcomes from these tests was calculated, using a threshold of $\alpha = 0.05$.

b. Bootstrapped samples - effect size

We determined the likelihood of finding the same or larger effect size as Cohn et al. The effect size was measured by the difference in means of average reported coin tosses (0.655) as per the main study the original paper.⁴

We drew samples of differing sizes with increments ranging from 50 to 1000. An even number of control and treatment subjects were drawn, in keeping with the random allocation of the experimental design. The samples were drawn with replacement from our Middle Eastern and Asia Pacific banker samples, separately. We calculated the mean reported coin tosses for the control and treatment groups and compared the difference in means to that of Cohn et al. Repeating this 10,000 times, we determined what proportion of simulated samples generated a difference in means from the control and treatment groups greater than or equal to that of the main study in Cohn et al.

c. Recruitment

In undertaking recruitment efforts, only medium to large-sized banks were approached to participate in the study, in an effort to improve the power of the experiment by obtaining a larger sample size than Cohn et al. Given the nature of field studies, it was not, however, possible to pre-specify the sample size of the study, given that participation and completion rates of the survey are unknown prior to running the study.

3.2. Funding

All studies were fully-funded by Google ATAP, via the London School of Economics and Political Science.

4. Human Subjects Approval

Approval for the study was provided by the Research Ethics Committee of the London School of Economics and Political Science (Reference Number 000582).

5. Additional References

53. Becker, G. S. (1974). Crime and Punishment: An Economic Approach. In W. M. Landes (Ed.), *Essays in the Economics of Crime and Punishment* (pp. 1-54). UMI (National Bureau of Economic Research)

54. Mazar, N., Amir, O., & Ariely, D. (2008). The Dishonesty of Honest People: A Theory of Self Concept Maintenance. *Journal of Marketing Research* , **45** (6), 633-644.

55. Bicchieri, C. (2006), *The Grammar of Society: The Nature and Dynamics of Social Norms*, Cambridge University Press.

56. Ross, L., Greene, D., & House, P. (1977). The “false consensus effect”: An egocentric bias in social perception and attribution processes. *Journal of experimental social psychology*, *13*(3), 279-301.

57. Koopman, J., Howe, M., Johnson, R. E., Tan, J. A., & Chang, C. H. (2013). A framework for developing word fragment completion tasks. *Human resource management review*, *23*(3), 242-253.

58. Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon’s Mechanical Turk. *Behavior research methods*, *44*(1), 1-23.