Banking culture revisited

Zoe Rahwan[a,b,c], Erez Yoeli[d], Barbara Fasolo[b]

[a] Centre for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany
Email: zrahwan@mpib-berlin.mpg.de
[b] Department of Management, London School of Economics and Political Science, Houghton Street, London, WC2A 2AE, United Kingdom
[c] Harvard Kennedy School, Harvard University, 79 John F. Kennedy Street, Cambridge, 02138, USA
[d] Sloan School of Management, MIT, 30 Memorial Dr, Cambridge, MA 02142, USA

The social sciences are going through what has been described as a 'reproducibility crisis.'[1,2] Highly influential findings derived from accessible populations, such as laboratories and crowd-sourced worker platforms, are not always replicated. Less attention has been given to replicating findings derived from inaccessible populations, and indeed recent high profile replication attempts explicitly excluded such populations.[3] Pioneering experimental work[4] offered a rare glimpse into banker culture and finds that bankers, unlike other professionals, are more dishonest when they think about their job. Given the importance of the banking sector, before the academy or policy-makers rely on these findings as an accurate diagnosis of banking culture, an exploration of their generalizability is warranted. Here we conduct the same incentivised task in five different populations, across three continents with 1,282 participants. In two studies (n=148, n=620), we observe some, though not significantly elevated, dishonesty among bankers primed to think about their work. We also find that having non-banking professionals (n=67, n=205, n=242) think about their job does not have a significant effect on honesty. We explore sampling and methodological differences to explain the variation in findings in relation to bankers and identify two key points: (i) general populations' relative expectations of banker behaviour vary across jurisdictions, suggesting that banking culture in the original jurisdiction may not found consistently across countries and (ii) having approached 27 financial institutions, many of which expressed concerns of adverse findings, we expect that only banks with a sound culture participated in our study. The latter introduces a substantial threat of selection bias that may undermine the generalizability of any similar field study. More broadly, our work highlights the complexity of undertaking a high-fidelity replication of sensitive, highly publicized field work with largely inaccessible populations due to institutional and geographic barriers. For policy-makers, this work suggests that caution should be exercised in generalizing findings to their national jurisdiction.

Recent seminal experimental work[4] finds evidence that when investment bankers are reminded of their professional identity (treatment), they become more dishonest than their colleagues who are asked to think about leisure activities (control). No such effect is found when priming professional identity among non-banking professionals. Together, these results were interpreted as providing evidence that banking identity was associated with weaker honesty norms and were cited widely, both in the academic literature[5,6,7,8,] and the popular press.[9,10,11]

In this experimental paradigm, honesty is measured via a simple coin-flip task. After answering questions about either their professional identity or leisure activities, bankers are asked to report

the outcomes from ten flips and are paid ~USD20 for each reported win. Without cheating, bankers should report winning coin tosses 50% of the time, on average, and the variation in reported wins should be characterized by a binomial distribution. Of course, it is impossible to infer how much more bankers cheat in real-world large-stakes, though there are positive correlations between honesty experiments and real world outcomes.[12,13,14,15]

The presence of weaker honesty norms in banking culture has significant negative implications for society collectively, as demonstrated by the role that dishonesty played in the sub-prime mortgage crisis.[16,17,18] This concern remains current: Since the original study[4] was conducted, further banking scandals have emerged in both investment and commercial segments,[19,20,21,22,23,24,25] trust in banking professionals and banks remains at relatively low levels,[26,27] and policy-makers remain concerned about culture in the banking industry.[28,29] In light of this concern, as well as on-going concerns regarding reproducibility of experimental results in the social sciences[2,3,30,31,32,33] and limited representativeness regarding national cultures,[34] we explore the generalisability of this influential field study of bankers.

In our first study, we employed Cohn et al's[4] design (n=128 with a follow-up study of n=80) on a significantly larger sample of 620 commercial (not investment) bankers at a large bank in the Asia Pacific (AP). Bankers in the treatment group tended to be more likely to cheat than the control group, though the difference was smaller than the original study's (Cohen's d^{35}=.06 cf. 0.37) and not statistically significant at conventional levels (54.0% vs. 53.0%, p=0.111 cf. 58.2% vs. 51.6%, p=0.017 (one-sided rank-sum tests, as are all subsequent results unless stated otherwise), Extended Data Fig.1, Extended Data Table 1A). Among a sample of full-time and part-time non-banking employees (n=242) which we strove to be nationally representative for gender and age, we found, similar to Cohn et al, that those primed with professional identity were no more likely to cheat than their non-primed counterparts (58.5% vs. 54.8%, p=0.114 cf. 55.8% vs. 59.8%, p=0.936, Extended Data Fig. 1, Extended Data Table 2A). Notably, the direction of the AP non-banker effect was consistent with those of the banker samples and contradictory to the direction of non-banker's effect in the original study). Further, the effect size of treatment on AP non-bankers (Cohen's d=0.19) was larger than that for bankers within the same jurisdiction.

In our second study, we employed the same study design on commercial bankers at a medium-sized bank in the Middle East (ME, n=148). Bankers in the treatment tended to be more likely to cheat, though similar to the AP bankers, the effect was very small (Cohen's d=0.11) and not statistically significant (56.9% vs. 54.9%, p=0.261, Extended Data Fig. 2, Extended Data Table 1B). In a small sample (n=67) of financial services regulators (i.e. non-bankers) in the same region, we again found no treatment effect from priming professional identity (50.3% (treatment), 51.1% (control), p=0.472, Extended Data Fig. 2). The direction of the effect was aligned with that of the original study's non-bankers, though not AP non-bankers. In a larger sample of European non-bankers (n=205), who were also financial services regulators, we again found no effect from priming professional identity (52.2% (treatment), 52.6% (control), p=0.572, Extended Data Fig. 3, Extended Data Table 2B). We note that for the Middle Eastern and European non-bankers studies, we were constrained to use rewards for charity rather than

self, though similar to previous research,[36] we found in a separate study (n=1,179) that this likely did not affect honesty (SI 1.2 and Extended Data Fig. 4 and 5).

In summary, we do not find any treatment effect among the bankers in the Middle East and Asia Pacific, in contrast to the original main study[4], although the results trend in the same direction. Consistent with the original study[4], we find no effects on honesty from having non-banking professionals think about their jobs. Together, these findings reveal that inducing professional identity results in varying effects on honesty across professions and jurisdictions, both in direction and range ([-4.0,+ 6.6] percentage point difference in average winning outcomes).

We do find a detectable increase in dishonesty when bankers are reminded of their profession when pooling data from our studies with Cohn et al's[4] main study (p=0.018) and when pooling with their main and follow-up studies, (p=.008) (Fig. 1, see also Extended Data Tables 1C and 1D). The effects among all banker samples, while small, are all in the same direction. However, the original main study's finding did not replicate according to the conventional significance level of α=0.05, in the original follow-up study, in the ME and AP individual banker samples nor when we pooled these latter two samples (p=0.082). This suggests that the original main study's[4] finding is not generalizable beyond the original population sampled. Further, variation in the direction of effects among non-banker samples raises the question of how likely the differences found in the original studies between banker and non-bankers are likely to replicate, at least outside of the original jurisdiction.

While the above-reported findings suggest that there may not be a consistent, nor sizeable, adverse effect of priming banker professional identity on honesty, we nevertheless undertake exploratory analyses to understand how sampling, other methodological and statistical power reasons may explain why our results differ from those of the original study.[4]

Regarding sampling issues, one possibility for explaining the variation in results is that the culture amongst bankers in our studies is different from the culture amongst bankers in the undisclosed location of the original studies. This could be due to variations in national norms,[5] banker norms between countries, and/or "self selection" of bankers into differing industry segments (e.g. investment versus commercial banking). We find negligible differences between the presumed original jurisdiction (which cannot be disclosed for legal reasons) and our Asia Pacific jurisdiction. However, some differences between the original and Middle Eastern jurisdictions are identified, which may help to account for the elevated baseline dishonesty found among unprimed Middle Eastern bankers and consequently, the smaller effect size from inducing professional identity (SI 2.3.1).

On banking norms, we do not find evidence of behavioural differences of treated bankers across the three jurisdictions (SI 2.3.2), though we do identify differences in people's relative expectations of bankers - a potential indicator of national banking norm variation. Specifically, bankers in the AP jurisdiction are not expected by others in the same jurisdiction to be more or less honest than doctors, prison inmates and the general population (SI 2.2.6). This is in contrast to the jurisdiction in the original study[4] where bankers were perceived less honest than

doctors, tended to be less honest than the general population, and were indistinguishable from prison inmates (Fig. 2 and Extended Data Fig. 6).

Heterogeneity in banking culture across segments of the industry is also reasonable to expect, given the variance in remuneration structures, business models and clients. Indeed, there are perceptions of lesser honesty among investment bankers relative to commercial bankers.[37,38] Despite this, we find no evidence that investment banking culture is more aversive than that of commercial banks (SI 2.3.3).

Another source of variation in results we explored was the possible self-selection of more honest people into banking rather than non-banking jobs in the original jurisdiction. Among the original study's[4] unprimed participants, non-bankers engage in greater dishonesty than bankers (59.8% vs. 51.6%, p=0.002), in contrast to our findings in other jurisdictions (SI 2.3.2). Further, Cohn et al's unprimed bankers contrast with our banker samples in that they do not engage in statistically detectable dishonesty (SI 2.1.4). And this is despite the poor expectations of bankers relative to others in that jurisdiction. The absence of statistically detectable dishonesty among Cohn's untreated bankers - an indication of more honest people self-selecting into the industry - may contribute to the larger effect of priming being identified. And this in turn suggests, in the context of other banker studies with smaller effect sizes, that banking culture in the original jurisdiction may have a more aversive effect on honesty norms than elsewhere.

At an institutional level, a self-selection bias may assist in explaining variation in results across the banker studies.The concern of adverse findings among banks, appears to have biased participation towards banks more likely to have a sound culture. In total, we approached 27 financial institutions around the world (including 14 investment banks), and only two (commercial) banks accepted to participate. Highlighting the self-selection bias, one lawyer specializing in compliance with investment banking clients who assisted us with recruitment noted "...I am particularly disappointed because the main reluctance I have encountered is clearly a concern by different firms that the survey might identify weaknesses in their culture that they are worried might somehow be exposed." One possible way around this for future studies, especially on sensitive topics, would be for replications by the same researchers or other groups in the same or other jurisdictions, ahead of publication of the initial findings.

Publicity surrounding the original study[4] may also have affected participant responses in our study and any future studies that will follow. In the AP banker sample, we found that 30% of respondents reported familiarity with research on banker culture which used a similar survey, though this does not appear to account for the smaller effect size relative to the original study[4] (SI 2.1.5).

Beyond these sampling and familiarity concerns, we explored how a range of methodological differences with Cohn et al's study[4] affected variation in results.

One notable methodological difference regarded experimenter disclosure about the purpose of the experiment. Primarily driven by strict ethics rules governing our research, we used

'incomplete disclosure' in our studies rather than deception, as per the original study.[4,39] In a separate study, we find no statistical differences in winning outcomes in the same coin-flipping task between 'incomplete disclosure' (n=309) and deception (n=315) conditions (59.9% vs. 58.4%, p=0.188 (two-sided) - SI 2.4, Extended Data Fig. 4 and 5). This provides an indication that variation in experimenter disclosure may not account for differences in effect sizes, though it may not generalise to bankers in all sampled jurisdictions. Still, a meta-analysis of dishonesty experiments finds less dishonesty associated with experiments using deception, and no statistical differences with coin-flipping tasks in particular.[40]

In order to have rewards in line with the denominations of local currencies, we adjusted the reward amounts from USD20 per coin toss in the original study, to ~USD14 per coin toss in all four Asia Pacific and Middle Eastern studies. We have reason to believe that the lesser rewards did not have a large impact on the variation in results. We extended the AP study to randomise the opportunity to win a reward for oneself or for charity - effectively USD0 for self - and found no difference in dishonesty levels (SI 1.2, Extended Data Fig. 7 and 8). This complements a growing body of work finding that dishonesty is largely insensitive to rewards[36,41], including for extraordinarily large rewards.[42]

Some other factors which we could not control, and may account for differences in results include who within participating organizations sent out the invitations to complete the survey, how the survey was circulated and the contents of the consent form. Responses could be influenced by whether an invitation comes from a CEO - as was the case in the AP sample - or a Communications team - as in our ME sample, or via an alumni network - as is the case for Cohn et al's[4] follow-up study. Also, the sampling can be affected by the nature of the invitation, such as all staff being emailed or a posting on an internal corporate Web page. Further, the wording of the consent form is likely to differ as institutions have varying requirements. The effects of these differences on all the banker responses are indeterminable and highlight the difficulty of conducting tightly controlled replications of studies in the field.

Finally, the timing of the studies is another factor that may affect outcomes. Since the 2008 financial crisis, global standard-setters such as the Basel Committee for Banking Supervision have undertaken significant efforts, which cascade to national regulators, to deter untoward institutional and individual behaviour in the banking industry that can lead to financial instability and unfair outcomes. Assuming these efforts have been globally efficacious, one would expect a diminished effect size on banker populations studied subsequent to Cohn et al's[4] work.

Statistical power is another factor that may account for the variations among the different studies. In the social sciences, the 'replication crisis' has in part been driven by issues of inadequately-sized samples. [43,44,45,46,47] This problem is exacerbated when conducting experiments in the field where there is often a limited ability to access both institutions and individuals. Indeed, some replication efforts deliberately averted studies involving 'inaccessible' populations, such as bankers.[3] Here, we find that it is difficult to find a significant effect without accessing larger samples (Fig. 3a). Specifically, a sample of more than ~170 is required to achieve conventional power of 80% (SI 3.1).

One established consequence of under-powered novel studies is smaller effect sizes in subsequent replications.[3,32,33,44,45] The original main study[4] found a "small" (Cohen's d=0.37) effect of banker priming with an average of ~0.7 more winning coin flips reported for such bankers. Consistent with replications trends, we find smaller effects from priming in our studies; on average 0.1 and 0.2 winning coin flips reported by primed bankers in our AP and ME studies, respectively (Cohen's d=0.06 and 0.11). Further, bootstrapped simulations using samples of varying sizes, confirm a low likelihood of finding an effect size equal to or greater than the original study (Fig. 3b, SI 3.1).

Our studies highlight a number of acute challenges in replicating field studies at high fidelity with inaccessible populations on an important question - does banker culture undermine honesty? These new results suggest that the original finding, which led some to conclude that banking culture is "corrosive,"[48] do not generalise across countries, banking segments, individual institutions nor time. While we find a negligible influence of banker identity on dishonesty, especially compared to the original study, it is difficult to identify precisely the underlying causes. Our strongest evidence points to differences in national banking norms and a pronounced threat of only 'ethical' banks selecting to participate in such research - a threat that cannot be effectively countered after high profile press coverage of original research. Finally, we must highlight a plausible parsimonious explanation: the effect observed in the original study[4] may only have held in a very specific setting and point in time, and as such, does not generalize beyond the sample contained in their main study nor across time.

Irrespective of the precise sources of the differences observed in the various studies of banker honesty, our findings have broad implications for replicability and generalizability outside of commonly accessed experimental populations. While, theoretically, such variation could be better understood with large-scale testing within and across countries coordinated by banking regulators, practically, such an approach would invite gaming from institutions in an effort to protect their reputations. Instead of focusing on direct replications, we believe that new tools and methodologies are needed to measure aspects of professional culture, such as honesty. This will be critical to better understand and ultimately manage the related risks and benefits to society that stem from the banking industry.

References

1. Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence?. *Perspectives on Psychological Science*, *7*(6), 528-530.

2. Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716.

3. Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M., ... & Altmejd, A. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, *2*(9), 637.

4. Cohn, A., Fehr, E., & Maréchal, M. A. (2014). Business culture and dishonesty in the banking industry. Nature, 516, 86-89.

5. Gächter, S., & Schulz, J. F. (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature*.

6. Purzycki, B. G., Apicella, C., Atkinson, Q. D., Cohen, E., McNamara, R. A., Willard, A. K., ... & Henrich, J. (2016). Moralistic gods, supernatural punishment and the expansion of human sociality. *Nature*, *530*(7590), 327.

7. Zingales, L. (2015). Presidential address: Does finance benefit society?. *The Journal of Finance*, *70*(4), 1327-1363.

8. Benjamin, D. J., Choi, J. J., & Fisher, G. W. (2010). *Religious identity and economic behavior* (No. w15925). National Bureau of Economic Research.

9. Lying, cheating bankers, http://www.economist.com/news/finance-and-economics/21633882-talking-about-their-work-makes-bankers-more-dishonest-lying-cheating-bankers (accessed 16 June 2016), *The Economist*, 22 November 2014.

10. Cookson, C. (2014), Bankers have tendency to lie for financial gain, say scientists, *Financial Times,* 19 November 2014.

11. Goodley, S. (2014), Coining it in: banking industry culture promotes dishonesty, research finds, *The Guardian,* 19 November 2014.

12. Halevy, R., Shalvi, S., & Verschuere, B. (2014). Being honest about dishonesty: Correlating self-reports and actual lying. *Human Communication Research*, *40*(1), 54-72.

13. Dai, Z., Galeotti, F., & Villeval, M. C. (2017). Cheating in the lab predicts fraud in the field: An experiment in public transportation. *Management Science*, *64*(3), 1081-1100.

14. Kröll, M., & Rustagi, D. (2016). *Shades of dishonesty and cheating in informal milk markets in India* (No. 134). SAFE Working Paper Series.

15. Potters, J., & Stoop, J. (2016). Do cheaters in the lab also cheat in the field?. *European Economic Review*, *87*, 26-33.

16. Bank of America to Pay $16.65 Billion in Historic Justice Department Settlement for Financial Fraud Leading up to and During the Financial Crisis https://www.justice.gov/opa/pr/bank-america-pay-1665-billion-historic-justice-department-settlement-financial-fraud-leading (accessed 16 June 2016), *The United States Department of Justice,* 21 August 2014.

17. SEC Enforcement Actions Addressing Misconduct That Led To or Arose from the Financial Crisis https://www.sec.gov/spotlight/enf-actions-fc.shtml (accessed 16 June 2016), *Securities and Exchange Commission*,

18. Wall Street and the Financial Crisis: Anatomy of a Financial Collapse,*United States Senate Permanent Subcommittee on Investigations,* 13 April 2011. [https://www.hsgac.senate.gov//imo/media/doc/Financial_Crisis/FinancialCrisisReport.pdf?attempt=2]

19. Chon, G., Binham, C. and Noonan, L., Six banks fined $5.6bn over rigging of foreign exchange markets https://next.ft.com/content/23fa681c-fe73-11e4-be9f-00144feabdc0 (accessed 16 June 2016), *Financial Times,* 20 May 2015

20. Capital punishment http://www.economist.com/news/finance-and-economics/21606321-frances-largest-bank-gets-fined-evading-american-sanctions-capital-punishment (accessed 16 June 2016), *The Economist,* 5 July 2014

21. Wood, R.W., New UBS Tax Evasion Probe, Again Over Americans http://www.forbes.com/sites/robertwood/2015/02/05/new-ubs-tax-evasion-probe-again-over-americans/#10c05d732b86 (accessed 16 June 2016), *Forbes*, 5 February 2015

22. Scannell, K., Arnold, M. and Stacey, K., HSBC forex traders charged with criminal fraud https://next.ft.com/content/9c44fe7a-4e97-11e6-8172-e39ecd3b86fc (accessed 24 July 2016), *Financial Times,* 20 July 2016

23. Gray, A., McLannahan, B. and Foley, S., Fake accounts cast cloud over Wells Fargo culture https://www.ft.com/content/c1b95e38-7694-11e6-b60a-de4532d5ea35 (accessed 9 September 2018), *Financial Times,* 9 September 2016.

24. Samson, A., Wells Fargo hit after auto insurance scandal emerges https://www.ft.com/content/2c8d60cf-b63f-329f-b02e-05f24176237c (accessed 9 September 2018), *Financial Times,* 28 July 2017.

25. Karp, P., Evershed, N. and Knaus, C., A recent history of Australia's banking scandals https://www.theguardian.com/australia-news/ng-interactive/2018/apr/19/a-recent-history-of-australias-banking-scandals (accessed 9 September 2018), *The Guardian,* 19 April 2018.

26. Honesty/Ethics in Professions http://www.gallup.com/poll/1654/honesty-ethics-professions.aspx (accessed 16 June 2016), *Gallup*.

27. McCarthy, J., Americans' Confidence in Banks Still Languishing Below 30% http://www.gallup.com/poll/192719/americans-confidence-banks-languishing-below.aspx?g_source=Economy&g_medium=newsfeed&g_campaign=tiles (accessed 16 June 2016), *Gallup,* 16 June 2016.

28. Our Business Plan 2016/17 https://www.fca.org.uk/news/our-business-plan-2016-17 (accessed 16 June 2016), *Financial Conduct Authority* (UK), 5 April 2016

29. ASIC's Corporate Plan 2015-2016 to 2018-2019 https://www.asic.gov.au/about-asic/what-we-do/our-role/asics-corporate-plan-2015-2016-to-2018-2019/ (accessed 16 June 2016), *Australian Securities and Investment Commission*, 31 August 2015.

30. Yong, E., (2012), Replication studies: Bad copy, *Nature News*, Nature Publishing Group, 16 May 2012.

31. Klein, R.A. et al., (2014), Data from Investigating Variation in Replicability: A "Many Labs" Replication Project, *Journal of Open Psychology Data*, 2(1).

32. Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams Jr, R. B., Alper, S., ... & Batra, R. (2018). Many Labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science*, *1*(4), 443-490.

33. Camerer, Colin F., et al. (2016), Evaluating replicability of laboratory experiments in economics, *Science*, 351.

34. Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioural and Brain Sciences*, 33, 61-135.

35. Cohen, J. (1969) Statistical Power Analysis for the Behavioral Sciences. NY: Academic Press.

36. Gino, F., Ayal, S., & Ariely, D. (2013). Self-serving altruism? The lure of unethical actions that benefit others. *Journal of economic behavior & organization*, *93*, 285-292.

37. Public Trust in Banking http://cdn.yougov.com/cumulus_uploads/document/ylf7gpof19/Public_Trust_in_Banking_Final.pdf (accessed 1 August 2016), *YouGov-Cambridge*, April 2013.

38. The Deloitte Bank Survey. Culture in banking - Under the microscope. (2013).

39. Sieber, J. E., Iannuzzo, R., & Rodriguez, B. (1995). Deception methods in psychology: Have they changed in 23 years?. *Ethics & Behavior*, *5*(1), 67-85.

40. Gerlach, P., Teodorescu, K., & Hertwig, R. (2019). The truth about lies: A meta-analysis on dishonest behavior. *Psychological bulletin*, *145*(1), 1.

41. Abeler, J., Nosenzo, D. and Raymond, D. (2016). Preferences for Truth-Telling, IZA Discussion Paper 10188 (*conditionally accepted at Econometrica*)

42. Rahwan, Z., Hauser, O. P., Kochanowska, E., & Fasolo, B. (2018). High stakes: A little more cheating, a lot less charity. *Journal of Economic Behavior & Organization, 152*, 276-295.

43. Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Du Sert, N. P., ... & Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 0021.

44. Ioannidis, J. P. (2008). Why most discovered true associations are inflated. *Epidemiology*, 640-648.

45. Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, *2*(8), e124.

46. Lane, D. M., & Dunlap, W. P. (1978). Estimating effect size: Bias resulting from the significance criterion in editorial decisions. *British Journal of Mathematical and Statistical Psychology*, *31*(2), 107-112.

47. Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean?. *American Psychologist*, *70*(6), 487.

48. Mohan, G., Banking industry culture primes for cheating, study suggests, http://www.latimes.com/science/sciencenow/la-sci-sn-cheating-bankers-20141119-story.html (accessed 16 June 2016), *Los Angeles Times*, 21 November 2014.

METHODS

Experimental Design for Bankers

Bankers were recruited from two institutions, one each in the Asia-Pacific (n=620) and Middle East (n=148). The protocol was based on Cohn et al's[4] experimental design and the studies were run in February 2016 and August 2015 respectively.

In summary, each institution invited their staff to participate in an on-line survey, assuring the confidentiality of their responses. Informed consent was sought from all participants. Once informed consent was granted, participants were randomly allocated to the treatment or control condition. In the control condition, participants were asked questions about their leisure activities before undertaking the coin-flip task. Under the treatment condition, participants were asked questions about their profession to activate or 'prime' their professional identity ahead of the task (e.g. Why did you decide to become a bank employee?).

Ahead of undertaking the coin tossing exercise, participants were informed about the reward mechanism. Specifically, they could win ≈USD 14 per each of the ten coin tosses, making a total maximum potential earning of ≈USD 140 in total in the local currency equivalent. Ahead of each coin toss, participants were made aware of the winning outcome. This provided an opportunity for dishonesty.

Recall, the experimental design enables comparisons between the reported winning outcomes of the control and treatment groups, and between each of these groups and a binomial distribution (p=0.5). A binomial distribution represents the frequencies of different outcomes that would be expected to emerge for an unbiased or 'fair' coin (i.e. 0.5 probability of tossing 'heads' and 'tails' alike). While it is impossible to know whether individual participants cheated during the experiment, as they are unobserved, by analysing the aggregated outcomes from both the treatment and control groups, an assessment of dishonesty can be made.[49]

Participants were informed that rewards would be calculated by (i) determining whether their total number of self-reported winning tosses was greater than another randomly drawn participant in the same survey, and if so, (ii) entering the qualifying participant into a draw in which one in five would win an amount corresponding to the number of winning tosses reported.

The first element introduces a competitive aspect to the reporting of winning coin flips. That is, if a participant expects that their colleagues will over-report the number of winning tosses they actually experienced, then that participant may be induced to over-report their number of winning tosses to increase the likelihood of receiving a reward. The second element of the reward mechanism – the lottery – was introduced consistent with the original study, to limit the cost of funding the research and is not expected to affect behaviour.[50]

Bankers were informed that any reward would be paid out to them personally in the form of shopping vouchers for popular retail outlets in the respective jurisdictions. Note that in the AP sample, an additional 559 participants were able to win a reward for a charity affiliated with the bank, instead of a reward for themself. The charity was affiliated with the bank, so was expected to be familiar to AP bank staff.

Immediately after completing the coin-tossing task, participants were asked for their expectations regarding the number of winning tosses they expected their colleagues to report on average. This was intended as a direct measure of perceived dishonesty of bankers' peers.

Following the coin tossing exercise and measure of expectations of peer performance in that exercise, participants were asked to complete a mock investment task. The main purpose of the exercise is to draw attention away from the coin tossing exercise and to facilitate assessment of whether individuals' risk appetite correlates with the frequency of winning outcomes.

Subsequently, participants were asked to complete a word quiz, in which they had to choose letters to complete words (e.g. _ _ o c k -> 'clock' or 'stock'). The purpose of this task is to determine whether the priming of professional identity was successful by comparing the number of banking-themed solutions among the control and treatment groups.

Participants then completed questions relating to their various work-related attitudes. The responses were made on a seven-point scale ranging from 'strongly disagree' to 'strongly agree.' The purpose of these questions is to understand participants' relative importance of materialism, competitiveness, and self-esteem being determined by others, in general, and with regard to their profession in particular. These assessments are used to explore possible mechanisms for dishonest behaviour.

For the sake of brevity of the survey and given that Cohn et al found no significant relationship between risk literacy and frequency of winning tosses – the risk literacy task was removed. Finally, participants were asked a range of questions to obtain personal information (i.e. age, gender, education, nationality, experience, income relative to colleagues, type of role and location of role in the bank). For the Asia Pacific study, the question of nationality was excluded as we were advised by the bank that it would be an anomaly for a non-national to be employed by the bank. We did however add a familiarity check, asking participants: "Are you familiar with research on banking industry culture which uses a survey like the one you just completed?"

Experimental Design for Non-Banking Professionals

The protocol for all three non-banking professional studies was based on Cohn et al's[4] experimental design for non-banking professionals (see Cohn et al's SI 4.3). This protocol, in turn, closely follows that used for banking professionals. The key differences broadly pertain to amending references to bankers to encapsulate all types of professionals, and for the Middle Eastern and European samples, amending the manipulation check to refer to financial services regulators (SI 2.2.3). Similar to the original study, we also excluded the risk literacy task.

      a. Asia Pacific

In August 2018, non-banking professionals (n=242) were recruited in the same AP jurisdiction as the bankers. The participants were sourced from a professional panel provider. They were screened to ensure that they were currently employed, in either a full-time or part-time capacity, and that they resided in the relevant jurisdiction. The panel provider used a balance sampling technique to recruit a nationally representative sample with regards to age and gender.

      b. Middle East

Non-banking professionals (n=67) were recruited from a financial services regulator, located in the same country as the ME bankers. The survey was conducted in September 2015.

c. Europe

Non-banking professionals (n=205) were recruited from a financial services regulator, located in Europe. The survey was conducted in January 2016.

Experimental Design for Experimenter Disclosure

This study is motivated by Cohn et al's use of deception and our use of incomplete disclosure.

In particular, we wanted to understand what influence variations in experimenter disclosure regarding the purpose of the study had on outcomes from the coin-flipping task. Due to the largely inaccessible nature of the banking population, we conducted an experiment (n=925) on Amazon Mechanical Turk (MTurk). We targeted the US general population as our jurisdictions of interest were understood to have insufficient active participants[51,52] from our previous recruitment experience and running prior surveys.

The London School of Economics Research Ethics Committee gave special dispensation for the use of deception. The survey was conducted on 6 July 2017.

The survey used was modelled on that used for both bankers and non-bankers as deployed by Cohn et al.[4] The manipulation related to the stated experimental purpose. On the landing page of the survey, participants were randomly allocated to (i) transparency (i.e. full disclosure), (ii) incomplete disclosure, or (iii) deception conditions. These conditions corresponded to being told in the introductory statement that it was a study on (i) Honesty, (ii) Norms and Attitudes among Professionals or (iii) Life and Satisfaction, respectively. The experimental purpose was repeated on the Consent Form. All subsequent elements of the survey were the same across the conditions.

After the introductory remarks, participants were requested to provide informed consent. In keeping with Cohn et al's[4] original banking survey, they were then asked to complete questions on life satisfaction and leisure activities. The questions on leisure activities formed the control condition in both the banker and non-banker surveys.

Participants were then introduced to the coin-flipping task, in which they could win US 5 cents for each winning outcome, marking a maximum of US 50 cents over the 10 rounds. While this is significantly less than winnings available to bankers and non-bankers, there is growing experimental evidence that the size of the reward has a negligible effect on cheating behaviour, including among the MTurk population from which we sampled.[41,42]

Unlike Cohn et al's original survey, there was no uncertainty over whether winnings would relate to either the coin-flipping or mock investment task outcomes, as the mock investment task was excluded for survey brevity. A word puzzle task, work attitudes and demographic questions followed the coin-flipping task. Finally, participants were asked about their experience with coin-flipping tasks and any perceptions of deception in this and other experiments.

Once the survey was closed, all participants were debriefed and for those experiencing deception, an apology was made.

Data Availability

The data generated for these studies and code used for analyses is available at: https://osf.io/56dcp/?view_only=3ef6585039b74bf9aae5deafa0f31e64.

Please note the code is written in R.

References

49. Houser, D., Vetter S., & Winter, J. (2012). Fairness and cheating, *European Economic Review*, **56** 1645-1655.

50. Starmer, C., & Sugden, R. (1991). Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation. *American Economic Review,* **81** (4), 971-978.

51. Ross, J., Zaldivar, A., Irani, L., & Tomlinson, B. (2009). Who are the turkers? worker demographics in amazon mechanical turk. *Department of Informatics, University of California, Irvine, USA, Tech. Rep.*

52. Difallah, D., Filatova, E., & Ipeirotis, P. (2018, February). Demographics and dynamics of mechanical turk workers. In *Proceedings of the eleventh acm international conference on web search and data mining* (pp. 135-143). ACM.

Main Article & Extended Data Legends & Titles

a. Main Article

[Insert Figure 1 here]

Figure 1 | Variation in how priming professional identity affects honesty.
a. Comparisons of the direction and size of effect from priming professional identity across separate samples of bankers around the world and the combination of all the bankers. While all studies find that making banker identity salient induced greater dishonesty only the original study (p=0.017, rank-sum test (one-sided); n=128, Cohen's d = 0.37) and the pooling of all bankers(p=0.008, rank-sum test (one-sided); n=976, Cohen's d = 0.127), finds a significant yet 'negligibly-sized' treatment effect. This effect is not found in the follow-up study (n=80) from the original paper (p=0.097, rank-sum test (one-sided); n=80, Cohen's d = 0.26), among the individual samples of Asia Pacific bankers (p=0.111, rank-sum test (one-sided); n=620, (Cohen's d = 0.06) and Middle Eastern bankers (p=0.261, rank-sum test (one-sided); n=148, Cohen's d = 0.11). Error bars reflect s.e.m. **b.** Comparisons of the size of effect from priming professional identity across separate samples of non-bankers around the world and the combination of all the non-bankers. Making professional identity salient induced different directions of effects among individual samples of non-bankers from the original study, Asia Pacific, Middle East and Europe, none of which were statistically significant (p=0.128 (n=133), p=0.114 (n=242), p=0.472 (n=67), p=0.572 (n=205), respectively; rank-sum tests (one-sided)). Error bars reflect s.e.m.

[Insert Figure 2 here]

Figure 2 | Relative Expectations of Banker Behaviour within and across Jurisdictions. a. The original study (n=183) provided evidence that expectations of banker behaviour (n=48) in the coin-flipping task are indistinguishable (p=0.558, two-sided rank-sum test) from prison inmates (n=45). Further, bankers were perceived to be less honest than doctors (n=44, p=0.005, two-sided rank-sum test) and tended to be perceived as less honest than the general population (n=46, p=0.080, two-sided rank-sum test). Expectations were sourced in a convenience sample of visitors to a Municipal Office. Only men were included in the sample reported. Error bars reflect s.e.m. **b.** In the Asia Pacific jurisdiction, using the same two-sided rank-sum tests we found no statistical differences between expectations of banker behaviour (n=65) and other groups assessed in the original study - prison inmates (n=64), the general population (n=58) and medical doctors (n=55). This suggests that relative expectations of bankers vary between jurisdictions. Differences in sampling may also account for the variation in findings. In contrast to the original study[4], we sourced a nationally representative panel, by age and gender. Error bars reflect s.e.m.

[Insert Figure 3 here]

Figure 3 | Replicability of the original study. a. The original main study (n=128) has an estimated power of ~68% bootstrapped samples subject to Wilcoxon rank-sum tests (one-sided). That is, of 10,000 simulated samples, ~68% found a statistically significant result given α = 0.05. This is below the 80% level of power that is conventionally targeted - a level that would require ~170 participants. The chart also marks the sample sizes of the original follow-up study (n=80), Asia Pacific (n=620) and Middle Eastern bankers (n=148) on the power curve. These samples would have a power of ~99% and ~75% respectively based off bootstrapped samples from the original main study. This suggests that Asia Pacific sample is adequately powered to detect a treatment effect at conventional levels though the Middle East sample is not. **b.** The original study found a difference of 0.7 average coin flips between the control and treatment groups. Drawing simulated samples of various sizes from Asia Pacific and Middle Eastern bankers reveals a low likelihood of finding the same or larger effect size than the original study (0.01%, 5.36% and 0.02%, respectively).

b. Extended Data

[Insert ED Figure 1 here]

Extended Data Figure 1 | Distributions of Reported Winning Coin Tosses in the Asia Pacific. The frequency of different totals of reported winning coin tosses among Asia Pacific individuals from 10 rounds of the coin-tossing task. **a.** Individuals in the treatment group (n=286) were primed with their banking professional identity. The control group (n=334) were asked a series of questions about their leisure activities. **b.** The treatment group (n=117) were primed with their professional identity. The control group (n=125) were asked the same series of questions on leisure activities as the bankers. Individuals reporting to be currently in banking roles were excluded.

[Insert ED Figure 2 here]

Extended Data Figure 2 | Distributions of Reported Winning Coin Tosses in the Middle East. The frequency of different totals of reported winning coin tosses among Middle Eastern individuals from 10 rounds of the coin-tossing task. **a.** Individuals in the treatment group (n=71) were primed with their banking professional identity. The control group (n=77) were asked a series of questions about their leisure activities. **b.** The treatment group (n=29) were primed with their financial services regulatory identity. The control group (n=38) were asked the same series of questions on leisure activities as the bankers.

[Insert ED Figure 3 here]

Extended Data Figure 3 | Distributions of Reported Winning Coin Tosses of Financial Services Regulators (Non-bankers) in Europe. The frequency of different totals of reported winning coin tosses among European regulators from 10 rounds of the coin-tossing task. Individuals in the treatment group (n=96) were primed with their financial services regulatory identity. The control group (n=109) were asked a series of questions about their leisure activities.

[Insert ED Figure 4 here]

Extended Data Figure 4 | Effect of Experimenter Disclosure on Honesty. This chart reports the average number of winning coin tosses out of 10 reported by participants in each of the three conditions (n=925) which varied the disclosed purpose of the experiment; deception ("Life and Satisfaction"), incomplete disclosure ("Norms and Attitudes among Professionals") and transparency ("Honesty"). The error bars refer to s.e.m. No differences were found between deception (M=5.84) - as used by Cohn et al - and incomplete disclosure (M=5.99) - as used in our studies (p=0.188, Wilcoxon rank-sum test (one-sided), alpha-adjusted for family-wise errors=0.05/3=0.017).The only statistical difference found between the conditions was that those in the incomplete disclosure condition reported a higher number of winning coin tosses than those in the transparency condition (M=5.64, p=0.006, Wilcoxon rank-sum test (one-sided)). The difference between the average outcomes of these two conditions was negligible and not sufficient to change actual payoffs for participants.

[Insert ED Figure 5 here]

Extended Data Figure 5 | Underlying Distributions of the Effect of Experimenter Disclosure on Honesty. This chart reports the frequency of different totals of reported winning coin tosses from 10 rounds of a coin-tossing task among MTurk participants. In this experiment the disclosed purpose of the experiment was randomly assigned to be one of three conditions; 'deception', wherein participants were informed they were in a study regarding "Life and Satisfaction" (n=315), 'incomplete disclosure' wherein participants were informed they were in a study regarding "Norms and Attitudes among Professionals (n=309)," and 'transparency', wherein participants were informed that we were studying "Honesty" (n=301).

[Insert ED Figure 6 here]

Extended Data Figure 6 | Underlying Distributions of Expectations of Others' Honesty. This chart reports the expected frequency of different totals of reported winning coin tosses from 10 rounds of a coin-tossing task among a sample of Asia Pacific non-banking professionals, sourced from a panel. In this experiment, the participants themselves had experience of the coin-tossing task ahead of being questioned on their expectations of reported winning outcome from one of four different populations. As such, participants had the opportunity to learn that one could be dishonest in the task. Participants were randomly assigned to be asked expectations of reported winning outcomes of the following populations; bankers (n=65), general population (n=58), medical doctors (n=55) and prison inmates (n=64).

[Insert ED Figure 7 here]

Extended Data Figure 7 | Effect of Nature of Reward on Honesty. This chart reports the average number of winning coin tosses out of 10 reported by Asia Pacific bankers in conditions where they can either win money for themselves (n=620) or charity (n=559). The error bars refer to s.e.m. On average, those winning money for themselves and for charity reported 5.34 and 5.17 winning tosses, respectively. No difference was found between those able to win money - up to ~USD 140 - for themselves or charity (p=0.073, Wilcoxon rank-sum test (two-sided)).

[Insert ED Figure 8 here]

Extended Data Figure 8 | Underlying Distributions of the Effect of Nature of Reward on Honesty. This chart reports the frequency of different totals of reported winning coin tosses from 10 rounds of a coin-tossing task among Asia Pacific bankers. Participants were able to either win a reward for a charity (n=559) or for themselves (n=620).

Extended Data Table 1 | Effect of professional identity on honesty among bank employees

[Insert ED Table 1A here]

1A. Asia Pacific Probit Estimates. The dependent variable is a reported winning toss. The reported results are marginal effects calculated at the median levels of the covariates, and the standard errors (in parentheses) have been corrected for clustering at the individual level. The median covariates are a measure of the change in probability of reporting a winning outcome. The models reported are as per those in Cohn et al. **a.** Reported winning tosses are regressed upon a dummy for the professional identity condition and individual characteristics (n=620). **b.** This model extends model a to include work-related variables (n=620). **c.** This model extends model b to include an additional control of self-reported materialism (n=620). Significance levels: *p<0.10, **p<0.05, ***p<0.01 (two-sided Wald tests).

[Insert ED Table 1B here]

1B. Middle East Probit Estimates. The dependent variable is a reported winning toss. The reported results are marginal effects calculated at the median levels of the covariates, and the standard errors (in parentheses) have been corrected for clustering at the individual level. The median covariates are a measure of the change in probability of reporting a winning outcome. The models reported are as per those in Cohn et al. **a.** Reported winning tosses are regressed upon a dummy for the professional identity condition and individual characteristics (n=148). **b.** This model extends model a to include work-related variables (n=148). **c.** This model extends model b to include an additional control of self-reported materialism (n=148). Significance levels: *p<0.10, **p<0.05, ***p<0.01 (two-sided Wald tests).

[Insert ED Table 1C here]

1C. Pooled Asia Pacific, Middle Eastern and Original study (main study) Probit Estimates. The dependent variable is a reported winning toss. The reported results are marginal effects calculated at the median levels of the covariates, and the standard errors (in parentheses) have been corrected for clustering at the individual level. The median covariates are a measure of the change in probability of reporting a winning outcome. The models reported are as per those in Cohn et al run on participants from their main study (n=128), the Asia Pacific banker study (n=620) and the Middle East banker study (n=148). **a.** Reported winning tosses are regressed upon a dummy for the professional identity condition and individual characteristics (n=896). **b.** This model extends model a to include work-related variables (n=896). **c.** This model extends model b to include an additional control of self-reported materialism (n=896). Significance levels: *p<0.10, **p<0.05, ***p<0.01 (two-sided Wald tests).

[Insert ED Table 1D here]

1D. Pooled Asia Pacific and Original study (main study) Probit Estimates. The dependent variable is a reported winning toss. The reported results are marginal effects calculated at the median levels of the covariates, and the standard errors (in parentheses) have been corrected for clustering at the individual level. The median covariates are a measure of the change in probability of reporting a winning outcome. The models reported are as per those in Cohn et al run on participants from their main study (n=128) and the Asia Pacific banker study (n=620). **a.** Reported winning tosses are regressed upon a dummy for the professional identity condition and individual characteristics (n=748). **b.** This model extends model a to include work-related variables (n=748). **c.** This model extends model b to include an additional control of self-reported materialism (n=748). Significance levels: *p<0.10, **p<0.05, ***p<0.01 (two-sided Wald tests).

Extended Data Table 2 | Effect of professional identity/banking condition on honesty in bank and non-banking employees

[Insert ED Table 2A here]

2A. Asia Pacific Probit Estimates. The dependent variable is a reported winning toss. The reported results are marginal effects calculated at the median levels of the covariates, and the standard errors (in parentheses) have been corrected for clustering at the individual level. The median covariates are a measure of the change in probability of reporting a winning outcome.The model is per Cohn et al run on participants from their main study (n=128) and non-banker population (n=133) to demonstrate that priming professional identity led to higher levels of reported winning tosses relative to non-bankers. Reported winning tosses are regressed upon a dummy for the professional identity condition and individual characteristics, and an interaction term for professional identity and bank employees (n= 620 (bankers) + 242 (non-bankers) = 862). Significance levels: *p<0.10, **p<0.05, ***p<0.01 (two-sided Wald tests).

Extended Data Table 2B

[Insert ED Table 2B here]

2B. Middle East Probit Estimates. The dependent variable is a reported winning toss. The reported results are marginal effects calculated at the median levels of the covariates, and the standard errors (in parentheses) have been corrected for clustering at the individual level. The median covariates are a measure of the change in probability of reporting a winning outcome.The model is per Cohn et al run on participants from their main study (n=128) and non-banker population (n=133) to demonstrate that priming professional identity led to higher levels of reported winning tosses relative to non-bankers. Reported winning tosses are regressed upon a dummy for the professional identity condition and individual characteristics, and an interaction term for professional identity and bank employees (n=148 (bankers) + 67 (non-bankers =215). Significance levels: *p<0.10, **p<0.05, ***p<0.01 (two-sided Wald tests).