
ECONtribute
Discussion Paper No. 163

**Meta-Nudging Honesty: Past, Present, and
Future of the Research Frontier**

Eugen Dimant

Shaul Shalvi

May 2022

www.econtribute.de



The most recent version of the working paper can always be downloaded following [this link](#).

Meta-Nudging Honesty: Past, Present, and Future of the Research Frontier

Eugen Dimant^{a,b,**}, Shaul Shalvi^c

^aUniversity of Pennsylvania

^bCESiifo, Munich

^cUniversity of Amsterdam

Abstract

Achieving successful behavior change via nudging is hard. This is particularly true when choice architects attempt to change behavior that is collectively harmful but individually beneficial. In this paper, we review the state-of-the-art of the behavior change literature to assess both robust evidence on the motives for lying and promising interventions to curb lying. Existing literature points to combining simple behavioral interventions (e.g., norm-nudging) with interventions that contain pecuniary consequences (e.g., norm enforcement via punishment). In this context, we also discuss the idea of ‘meta-nudging’: rather than pursuing the classical approach to nudge targeted behavior *directly*, one may instead want to nudge behavior *indirectly* by targeting those who are in positions of power and have the ability to enforce norm adherence of others. Research suggests that delegating the enforcement of norm prescriptions can be a promising approach to nudge honesty.

Keywords: Behavior Change, Honesty, Lying, Nudging

Introduction

Research on nudging suggests that behavior change is difficult, often fails, and at times even backfires [1–4]. Evidently, this is not only the case when societal norms regarding the proper behavior are vague and contain wiggle-room [5], but also when norms are firmly established and followed by one’s peers [6–10]. Take for example, the norm of honest behavior, which

*This work was financially supported by the German Research Foundation (DFG) under Germany’s Excellence Strategy — EXC 2126/1– 390838866.

**Corresponding author

Email addresses: edimant@sas.upenn.edu (Eugen Dimant), s.shalvi@uva.nl (Shaul Shalvi)

is praised and socially desirable. Nonetheless, high-profile and systemic cases of dishonesty still persist (see, e.g., the recent Enron, Madoff, and Volkswagen scandals) [11]. Often, these instances are characterized by *collaborative dishonesty* in which the success of a dishonest act relies on successful coordination within groups [12–16].

In this article, we synthesize the research frontier at the intersection of honesty and behavior change. We focus on highlighting three aspects in particular:

1. How does cutting-edge research approach providing robust evidence that is useful to public policy?
2. What are the relevant individual, collective, and institutional factors that drive dishonesty, and why is nudging honesty not trivial?
3. What do we know about how to nudge honesty successfully, whom to target, and what pitfalls to avoid?

How does science inform public policy to reduce dishonesty?

For behavioral public policy to be effective and to have “bite”, the underlying evidence that informs the policies needs to be robust. To achieve this, recent trends in the academic community include:

- (a) the use of prediction markets that harness the forecasting ability of individuals to predict the replicability of existing and effectiveness of future interventions [17–20]
- (b) so-called ‘megastudies’ in which independent teams of scholars test different interventions to achieve behavior change [21–23]
- (c) meta-analytical evaluation of existing research, published and unpublished, to identify impact and robustness of interventions while also accounting for publication bias as much as possible [24, 25].

These approaches are not necessarily distinct and often end up blending, for example, interventions with prediction markets and forecastings [25–27]. This avenue has been particularly promising in the context of the replication movement [18]. Research on dishonesty can – and often already does – leverage these methodological insights [16, 28, 29].

In what follows, we will discuss this research in more detail and provide an outlook for the most promising approaches to nudge honesty.

Drivers of dishonesty

One line of research examining the mechanisms of and remedies for dishonesty has established that *individual* factors play an important role. These typically come in two forms: first, one’s ability to exploit moral wiggle-rooms via self-serving justification [30–33]. That is, individuals are able to abuse an existing moral wiggle-room by reinterpreting, distorting, or purposefully forgetting existing evidence favoring norms of honesty [5, 34–38]. Second is the purposely selective seeking and processing of relevant information, which allows individuals to remain ignorant and maintain plausible deniability [39–43]. This line of research emphasizes dishonesty as largely independent of others.

Another line of research emphasizes the role of *collective* factors in driving dishonesty. These include settings in which one finds justification for one’s own dishonesty in the dishonesty of peers [12, 15, 16]. The core insight here is that social reinforcement via observing and being observed by one’s peers is interpreted as a signal of the dominant social norm, which can accelerate the contagion of dishonesty [1, 8, 44, 45].

We can turn to meta-studies to better understand what robustly predicts dishonest behavior. A recent meta-analysis [16] on collaborative dishonesty analyzed 87,771 decisions across 21 behavioral tasks by 10,923 participants. Meta-analytical results revealed various factors that increase collaborative dishonesty, including higher financial incentives, conducting a lab experiment rather than a field experiment, and absence of negative externalities of one’s lies on others. Collaborative dishonesty is also higher in studies using no experimental deception, and when groups consist of younger individuals and more males. The latter is also consistent with previous research, suggesting that men are more dishonest than women [29, 46, 47]. Perhaps most relevant to our discussion here, however, is the finding the behavior of different group members is correlated when they interact repeatedly. That is, participants are more likely to lie when their partners lie, and lying increases as the task progresses. This is consistent with the findings that deviance breeds deviance [48, 49] and that norms of good behavior are quick to deteriorate over time [8].

Interestingly, some of the findings on collaborative dishonesty [16] also emerge in individual decision making settings. For example, the correlations suggest that older people are also more honest [28, 29], lack of experimental deception leads to higher levels of dishonesty [29], and lying is more pronounced in the lab than in the field [29]. That said, in individual settings, no effect was found for financial incentives on the level of dishonesty [28, 29]. Taken together, those three meta-studies inform our understanding on the settings

in which people may lie and conversely - when they would be most likely to act honestly.

From this, it is evident that the decision to engage in (dis)honest behavior is driven by an interplay of factors at both the individual and collective level. This, in turn, suggests that nudging honesty is not a trivial proposition. In what follows, we will unpack a few nudging approaches that have shown promise to tackle dishonesty successfully.

Challenges with and promising approaches for nudging honesty

Nudges can be a potent way to achieve behavior change [6, 50]. Importantly, however, nudges that target honesty are frequently unsuccessful and their effectiveness varies, thus highlighting that nudging honesty is not trivial. This is supported by empirical research, both in the lab and in the field, indicating that mitigating dishonest behavior through behavioral interventions alone, e.g. norm-nudges or defaults, is challenging [6, 9, 35, 50–53]. This is further augmented by theoretical arguments stressing that the *nudgeability* of individuals matters, which is often overlooked in existing research and deserves more scholarly attention moving forward [54]. That is, an individual’s predisposition towards the targeted behavior is a strong determinant of how effective an intervention ultimately is. Opposing attitudes can effectively render individuals ‘unnudgeable’, thus calling for more ‘forceful’ interventions (such as monetary incentives) – on top of the more light-touch behavioral interventions discussed above – to reduce dishonesty [9, 26, 47, 55, 56].

Arguably, mitigating dishonesty is more challenging at the collaborative level since the social settings involve a number of additional layers that can act as catalysts, such as peer pressure to engage in and fear of retaliation when not conforming to the norm of dishonesty behavior [57, 58]. Thus, behavioral interventions that show promise at the individual level may or may not translate to the more complex collaborative environment in which peer effects are at play. To achieve success, one may have to draw on behavioral interventions that target mechanisms on both the individual and collective level.

One such promising new approach has been coined ‘meta-nudging’ and suggests that one can also successfully nudge individuals *indirectly* by harnessing the power of social norms enforcement [56]. That is, by targeting those who enforce behavior – rather than those whose behavior one wants to alter – behavioral interventions would aim at nudging individuals in positions of power who have the ability to enforce the transgressors’ adherence to social norms. Arguably, behavioral interventions that rely on delegated policing (“hired gun”) might both be perceived less intrusive and more successful in that they would

capitalize on existing peer mechanisms [59] and can complement nudging directly at the individual level [21–23, 60–62]. Thus far, this approach has been mostly tested successfully in individual-decision environments [56, 63]. Investigating whether these interventions are also successful in collaborative environments that are characterized by social interactions remains an empirical question.

Conclusion

Behavior change is hard. This is even true when individuals are ‘nudgeable’ and have a pre-disposition that favors behaviors that one can generally agree on is largely beneficial, such as eating healthier. However, it is arguably even harder to try to change behavior that – even though it is detrimental on a collective level and potentially also violates existing social norms – is beneficial at the individual level. That is, a collective understanding of what is ‘right’ alone is insufficient in mitigating such behavior. In this paper, we have reviewed behavioral interventions that are at the forefront of successfully nudging honesty. Our main insight is that, while various approaches have shown promise, future research may want to put emphasis on developing a more fine-grained understanding of the interplay of motives for lying on both the individual and the collective level.

References

- [1] G. Bolton, E. Dimant, and U. Schmidt, “Observability and social image: On the robustness and fragility of reciprocity,” *Journal of Economic Behavior & Organization*, vol. 191, pp. 946–964, 2021.
- [2] R. H. Thaler and C. R. Sunstein, *Nudge: The final edition*. Penguin, 2021.
- [3] C. R. Sunstein, “The distributional effects of nudges,” *Nature Human Behaviour*, vol. 6, no. 1, pp. 9–10, 2022.
- [4] C. Morvinski, S. Saccardo, and O. Amir, “Mis-nudging morality,” *Management Science*, 2022.
- [5] C. Bicchieri, E. Dimant, and S. Sonderegger, “It’s not a lie if you believe the norm does not apply: Conditional norm-following with strategic beliefs,” Working Paper Available at SSRN: <https://dx.doi.org/10.2139/ssrn.3326146>, 2021.
- [6] C. Bicchieri and E. Dimant, “Nudging with care: The risks and benefits of social information,” *Public choice*, pp. 1–22, 2019.
- [7] C. Bicchieri, “Norm nudging: How to measure what we want to implement,” in *Behavioral Science in the Wild*, N. Mazar and D. Soman, Eds., University of Toronto Press, 2022, ch. 7.
- [8] C. Bicchieri, E. Dimant, S. Gächter, and D. Nosenzo, “Social proximity and the erosion of norm compliance,” *Games and Economic Behavior*, vol. 132, pp. 59–72, 2022.
- [9] E. Dimant, “Hate trumps love: The impact of political polarization on social preferences,” Working Paper Available at SSRN: <https://dx.doi.org/10.2139/ssrn.3680871>, 2022.
- [10] E. Dimant, M. Gelfand, A. Hochleitner, and S. Sonderegger, “Strategic behavior with tight, loose, and polarized norms,” Working Paper Available at SSRN: <https://bit.ly/3ryY3Pc>, 2022.
- [11] A. Cohn, E. Fehr, and M. A. Maréchal, “Business culture and dishonesty in the banking industry,” *Nature*, vol. 516, no. 7529, pp. 86–89, 2014.
- [12] O. Weisel and S. Shalvi, “The collaborative roots of corruption,” *Proceedings of the National Academy of Sciences*, vol. 112, no. 34, pp. 10 651–10 656, 2015.
- [13] I. Soraperra, O. Weisel, S. Kochavi, M. Leib, H. Shalev, *et al.*, “The bad consequences of teamwork,” *Economics Letters*, vol. 160, pp. 12–15, 2017.
- [14] M. G. Kocher, S. Schudy, and L. Spantig, “I lie? we lie! why? experimental evidence on a dishonesty shift in groups,” *Management Science*, vol. 64, no. 9, pp. 3995–4008, 2018.
- [15] O. Weisel and S. Shalvi, “Moral currencies: Explaining corrupt collaboration,” *Current opinion in psychology*, vol. 44, pp. 270–274, 2022.
- [16] M. Leib, N. Köbis, I. Soraperra, O. Weisel, and S. Shalvi, “Collaborative dishonesty: A meta-analysis,” Working Paper, 2022.
- [17] A. Dreber, T. Pfeiffer, J. Almenberg, S. Isaksson, B. Wilson, *et al.*, “Using prediction markets to estimate the reproducibility of scientific research,” *Proceedings of the National Academy of Sciences*, vol. 112, no. 50, pp. 15 343–15 347, 2015.
- [18] C. F. Camerer, A. Dreber, F. Holzmeister, T.-H. Ho, J. Huber, *et al.*, “Evaluating the replicability of social science experiments in nature and science between 2010 and 2015,” *Nature Human Behaviour*, vol. 2, no. 9, pp. 637–644, 2018.

- [19] S. DellaVigna and D. Pope, “What motivates effort? evidence and expert forecasts,” *The Review of Economic Studies*, vol. 85, no. 2, pp. 1029–1069, 2018.
- [20] S. DellaVigna, D. Pope, and E. Vivalt, “Predict science to improve science,” *Science*, vol. 366, no. 6464, pp. 428–429, 2019.
- [21] K. L. Milkman, M. S. Patel, L. Gandhi, H. N. Graci, D. M. Gromet, *et al.*, “A megastudy of text-based nudges encouraging patients to get vaccinated at an upcoming doctor’s appointment,” *Proceedings of the National Academy of Sciences*, vol. 118, no. 20, 2021.
- [22] K. L. Milkman, D. Gromet, H. Ho, J. S. Kay, T. W. Lee, *et al.*, “Megastudies improve the impact of applied behavioural science,” *Nature*, vol. 600, no. 7889, pp. 478–483, 2021.
- [23] K. L. Milkman, L. Gandhi, M. S. Patel, H. N. Graci, D. M. Gromet, *et al.*, “A 680,000-person megastudy of nudges to encourage vaccination in pharmacies,” *Proceedings of the National Academy of Sciences*, vol. 119, no. 6, e2115126119, 2022.
- [24] N. C. Köbis, B. Verschuere, Y. Bereby-Meyer, D. Rand, and S. Shalvi, “Intuitive honesty versus dishonesty: Meta-analytic evidence,” *Perspectives on Psychological Science*, vol. 14, no. 5, pp. 778–796, 2019.
- [25] S. DellaVigna and E. Linos, “Rcts to scale: Comprehensive evidence from two nudge units,” *Econometrica*, vol. 90, no. 1, pp. 81–116, 2022.
- [26] E. Dimant, D. Pieper, E. G. Clemente, A. Dreber, and M. J. Gelfand, “Politicizing mask-wearing: Predicting the success of behavioral interventions among republicans and democrats in the u.s.,” *Scientific Reports*, 2022.
- [27] M. Gelfand, R. Li, E. Stamkou, D. Pieper, E. Denison, *et al.*, “Persuading republicans and democrats to comply with mask wearing: An intervention tournament,” *Journal of Experimental Social Psychology*, 2022.
- [28] J. Abeler, D. Nosenzo, and C. Raymond, “Preferences for truth-telling,” *Econometrica*, vol. 87, no. 4, pp. 1115–1153, 2019.
- [29] P. Gerlach, K. Teodorescu, and R. Hertwig, “The truth about lies: A meta-analysis on dishonest behavior.,” *Psychological bulletin*, vol. 145, no. 1, p. 1, 2019.
- [30] S. Shalvi, J. Dana, M. J. Handgraaf, and C. K. De Dreu, “Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior,” *Organizational Behavior and Human Decision Processes*, vol. 115, no. 2, pp. 181–190, 2011.
- [31] S. Shalvi, F. Gino, R. Barkan, and S. Ayal, “Self-serving justifications: Doing wrong and feeling moral,” *Current Directions in Psychological Science*, vol. 24, no. 2, pp. 125–130, 2015.
- [32] R. Bénabou and J. Tirole, “Mindful economics: The production, consumption, and value of beliefs,” *Journal of Economic Perspectives*, vol. 30, no. 3, pp. 141–64, 2016.
- [33] F. Gino, M. I. Norton, and R. A. Weber, “Motivated bayesians: Feeling moral while acting egoistically,” *Journal of Economic Perspectives*, vol. 30, no. 3, pp. 189–212, 2016.
- [34] U. Gneezy, S. Saccardo, M. Serra-Garcia, and R. van Veldhuizen, “Bribing the self,” *Games and Economic Behavior*, vol. 120, pp. 311–324, 2020.
- [35] E. Dimant, A. v. K. Gerben, and S. Shalvi, “Requiem for a nudge: Framing effects in nudging honest,” *Journal of Economic Behavior & Organization*, vol. 172, pp. 247–266, 2020.

- [36] P. Schwardmann and J. Van der Weele, “Deception and self-deception,” *Nature human behaviour*, vol. 3, no. 10, pp. 1055–1061, 2019.
- [37] F. Galeotti, C. Saucet, and M. C. Villeval, “Unethical amnesia responds more to instrumental than to hedonic motives,” *Proceedings of the National Academy of Sciences*, vol. 117, no. 41, pp. 25 423–25 428, 2020.
- [38] S. Saccardo and M. Serra-Garcia, “Cognitive flexibility or moral commitment? evidence of anticipated belief distortion,” Working Paper, 2022.
- [39] R. Golman, D. Hagmann, and G. Loewenstein, “Information avoidance,” *Journal of Economic Literature*, vol. 55, no. 1, pp. 96–135, 2017.
- [40] Z. Grossman and J. J. Van der Weele, “Self-image and willful ignorance in social decisions,” *Journal of the European Economic Association*, vol. 15, no. 1, pp. 173–217, 2017.
- [41] T. Sharot and C. R. Sunstein, “How people decide what they want to know,” *Nature Human Behaviour*, vol. 4, no. 1, pp. 14–19, 2020.
- [42] E. Dimant, F. Galeotti, and M. C. Villeval, “Norm-formation and the role of information acquisition,” Mimeo, 2022.
- [43] L. Vu, I. Soraperra, M. Leib, J. van der Weele, and S. Shalvi, “Willful ignorance: A meta-analysis,” Working Paper, 2022.
- [44] E. Dimant, “Contagion of pro-and anti-social behavior among peers and the role of social proximity,” *Journal of Economic Psychology*, vol. 73, pp. 66–88, 2019.
- [45] Z. B. Ren, E. Dimant, and M. E. Schweitzer, “Social motives for sharing conspiracy theories,” Working Paper, 2022.
- [46] V. Capraro, “Gender differences in lying in sender-receiver games: A meta-analysis,” *Judgment and Decision Making*, vol. 13, no. 4, pp. 345–355, 2018.
- [47] J. Buckenmaier, E. Dimant, and L. Mittone, “Effects of institutional history and leniency on collusive corruption and tax evasion,” *Journal of Economic Behavior & Organization*, vol. 175, pp. 296–313, 2020.
- [48] E. Dimant and T. Schulte, “The nature of corruption: An interdisciplinary perspective,” *German Law Journal*, vol. 17, no. 1, pp. 53–72, 2016.
- [49] E. Dimant and G. Tosato, “Causes and effects of corruption: What has past decade’s empirical research taught us? a survey,” *Journal of Economic Surveys*, vol. 32, no. 2, pp. 335–356, 2018.
- [50] J. Beshears and H. Kosowsky, “Nudging: Progress to date and future directions,” *Organizational Behavior and Human Decision Processes*, vol. 161, pp. 3–19, 2020.
- [51] S. Boonmanunt, A. Kajackaite, and S. Meier, “Does poverty negate the impact of social norms on cheating?” *Games and Economic Behavior*, vol. 124, pp. 569–578, 2020.
- [52] C. Y. Zhang, J. Hemmeter, J. B. Kessler, R. D. Metcalfe, and R. Weathers, “Nudging timely wage reporting: Field experimental evidence from the united states social supplementary income program,” Working Paper, 2020.
- [53] J. B. Martuza, S. R. Skard, L. Løvlie, and H. Thorbjørnsen, “Do honesty-nudges really work? a large-scale field experiment in an insurance context,” *Journal of Consumer Behaviour*, 2022.

- [54] D. de Ridder, F. Kroese, and L. van Gestel, “Nudgeability: Mapping conditions of susceptibility to nudge influence,” *Perspectives on Psychological Science*, p. 1745691621995183, 2021.
- [55] P. Campos-Mercade, A. N. Meier, F. H. Schneider, S. Meier, D. Pope, *et al.*, “Monetary incentives increase covid-19 vaccinations,” *Science*, vol. 374, no. 6569, pp. 879–882, 2021.
- [56] E. Dimant and T. Gesche, “Nudging enforcers: How norm perceptions and motives for lying shape sanctions,” Working Paper Available at SSRN: <https://dx.doi.org/10.2139/ssrn.3664995>, 2021.
- [57] E. Reuben and M. Stephenson, “Nobody likes a rat: On the willingness to report lies and the consequences thereof,” *Journal of Economic Behavior & Organization*, vol. 93, pp. 384–391, 2013.
- [58] L. Charroin, B. Fortin, and M. C. Villeval, “Homophily, peer effects, and dishonesty,” Working Paper, 2021.
- [59] J. Andreoni and L. K. Gee, “Gun for hire: Delegated enforcement and peer punishment in public goods provision,” *Journal of Public Economics*, vol. 96, no. 11-12, pp. 1036–1046, 2012.
- [60] A. Schram and G. Charness, “Inducing social norms in laboratory allocation choices,” *Management Science*, vol. 61, no. 7, pp. 1531–1546, 2015.
- [61] V. Capraro, G. Jagfeld, R. Klein, M. Mul, and I. v. de Pol, “Increasing altruistic and cooperative behaviour with simple moral nudges,” *Scientific reports*, vol. 9, no. 1, pp. 1–11, 2019.
- [62] D. P. Daniels and J. J. Zlatev, “Choice architects reveal a bias toward positivity and certainty,” *Organizational Behavior and Human Decision Processes*, vol. 151, pp. 132–149, 2019.
- [63] J. J. Zlatev, D. P. Daniels, H. Kim, and M. A. Neale, “Default neglect in attempts at social influence,” *Proceedings of the National Academy of Sciences*, vol. 114, no. 52, pp. 13643–13648, 2017.