The science of honesty

A review and research agenda

Shalvi, S.; Levine, E.E.; Thielmann, I.; Jayawickreme, E.; Van Rooij, B.; Teodorescu, K.; Schurr, A.; Furr, R.M.; Aglioti, S.M.; Zettler, I.; Cohen, T.R.; Pittarello, A.; Barkan, R.; Köbis, N.; Leib, M.; Mitkidis, P.; Schulz, J.; Dimant, E.; van Kleef, G.A.; Ścigała, K.A.; Rilke, R.M.; Ayal, S.; Beersma, B.; Plonsky, O.; Hilbig, B.; Weisel, O.; Butera, F.; Feldman, Y.; Verschuere, B.; Zanetti, C.; Hochman, G.; Kret, M.E.; Peer, E.; Capraro, V.; Dorrough, A.R.; Speer, S.; Ritov, I.

DOI
10.1016/bs.aesp.2025.04.004

Publication date
2025

Document Version
Final published version

Published in
Advances in Experimental Social Psychology

CHAPTER FIVE

The science of honesty: A review and research agenda

Shaul Shalvi[a,*], Emma E. Levine[b], Isabel Thielmann[c], Eranda Jayawickreme[d,e], Benjamin Van Rooij[f], Kinneret Teodorescu[g], Amos Schurr[h,i], R. Michael Furr[j], Salvatore Maria Aglioti[k], Ingo Zettler[l], Taya R. Cohen[m], Andrea Pittarello[n], Rachel Barkan[o], Nils Köbis[p], Margarita Leib[q], Panagiotis Mitkidis[r], Jonathan Schulz[s], Eugen Dimant[t], Gerben A. van Kleef[u], Karolina Aleksandra Ścigała[v], Rainer Michael Rilke[w], Shahar Ayal[x], Bianca Beersma[y], Ori Plonsky[z], Benjamin Hilbig[aa], Ori Weisel[ab], Fabrizio Butera[ac], Yuval Feldman[ad], Bruno Verschuere[ae], Cinzia Zanetti[af], Guy Hochman[ag], Mariska E. Kret[ah], Eyal Peer[ai], Valerio Capraro[aj], Angela R. Dorrough[ak,al], Sebastian Speer[am], and Ilana Ritov[an]

[a]Amsterdam School of Economics, University of Amsterdam, Amsterdam, NH, The Netherlands
[b]Booth School of Business, The University of Chicago, Chicago, IL, United States
[c]Max Planck Institute for the Study of Crime, Security and Law, Freiburg, Germany
[d]Department of Psychology, Wake Forest University, Winston-Salem, NC, United States
[e]Program for Leadership and Character, Wake Forest University Winston-Salem, NC, United States
[f]Faculty of Law, University of Amsterdam, Amsterdam, ND, The Netherlands
[g]Faculty of Data and Decision Sciences, Technion, Israel Institute of Technology, Haifa, Israel
[h]Department of Business Administration, Ben-Gurion University of the Negev, Beersheba, Israel
[i]Rady School of Management, University of California San Diego, La Jolla, California, United States
[j]Department of Psychology, Wake Forest University, Winston-Salem, NC, United States
[k]Department of Psychology, Sapienza University of Rome, Rome, Italy
[l]Department of Psychology and Copenhagen Center for Social Data Science (SODAS), University of Copenhagen, Copenhagen, Denmark
[m]Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA, United States
[n]Department of Psychology, Stony Brook University, Stony Brook, NY, United States
[o]Department of Business Administration, Ben-Gurion University of the Negev, Beersheba, Israel
[p]Research Center Trustworthy Data Science and Security, University Duisburg-Essen, Duisburg, Germany
[q]Department of Psychology, Tilburg University, Tilburg, LE, The Netherlands
[r]Department of Management, Aarhus University, Aarhus V, Denmark
[s]Department of Economics, George Mason University, Fairfax, VA, United States
[t]Behavioral and Decision Sciences, University of Pennsylvania, Philadelphia, PA, United States
[u]Department of Psychology, University of Amsterdam, Amsterdam, NK, The Netherlands
[v]Department of Psychology and Behavioural Sciences, Aarhus University, Aarhus C, Denmark
[w]Economics Group, WHU - Otto Beisheim School of Management, Vallendar, Germany
[x]Department of Psychology, Reichman University, Herzliya, Israel
[y]Department of Organization Sciences, Vrije University Amsterdam, Amsterdam, The Netherlands
[z]Faculty of Data and Decision Sciences, Technion, Israel Institute of Technology, Haifa, Israel
[aa]Department of Psychology, University of Kaiserslautern-Landau, Landau, Germany
[ab]Department of Public Policy, Tel Aviv University, Tel Aviv, Israel
[ac]Institute of Psychology, University of Lausanne, Lausanne, Switzerland
[ad]Faculty of Law, Bar Ilan University, Ramat-Gan, Israel
[ae]Department of Psychology, University of Amsterdam, Amsterdam, NK, The Netherlands

Advances in Experimental Social Psychology, Volume 72
ISSN 0065-2601, https://doi.org/10.1016/bs.aesp.2025.04.004

[af]Institute of Psychology, University of Lausanne, Lausanne, Switzerland
[ag]Department of Psychology, Reichman University, Herzliya, Israel
[ah]Department of Psychology, Leiden University, Leiden, AK, The Netherlands
[ai]School of Public Policy and Governance, Hebrew University of Jerusalem, Jerusalem, Israel
[aj]Department of Psychology, University of Milan Bicocca, Milan, Italy
[ak]Department of Psychology, University of Hagen, Hagen, Germany
[al]University of Cologne, Cologne, Germany
[am]Department of Psychology, Peretsman-Scully Hall, Princeton University, Princeton, NJ, United States
[an]School of Education, Hebrew University of Jerusalem, Jerusalem, Israel
*Corresponding author: e-mail address: s.shalvi@uva.nl

Contents

Abstract

Honesty, defined as freedom from fraud or deception, is widely valued in many aspects of life, from personal relationships to professional settings. Yet acts of dishonesty remain widespread, including political and corporate scandals, misinformation, personal betrayal, and so on. Understanding honesty and the factors that influence it provides insights that are essential for fostering trust and combating corruption. In this review, we synthesize key findings from research on honesty, focusing on when people choose to be truthful or deceptive. We argue that although much is known about honesty in isolated, low-risk contexts, an urgent need exists to study honesty in more complex, realistic settings, such as those involving interpersonal relationships, potential sanctions, or group influences. Our proposed framework highlights understudied contexts and encourages future studies to explore settings where enforcement and social dynamics play a significant role in decision-making. To do so, we point out 66 open research questions that we find most promising to explore. By integrating insights from multiple disciplines, we aim to advance the understanding of honesty and provide a roadmap for research that can inform policies and interventions to promote integrity in society.

Honesty—defined as being free from fraud or deception (Merriam-Webster, 2024)—is arguably one of the most valued human qualities. In romantic relationships, friendships, workplaces, and organizations, people seek out genuine and trustworthy partners. These qualities foster collaboration and help build lasting relationships. As William Shakespeare aptly put it, "There is no inheritance so rich as honesty" (Shakespeare, 2019; All's Well That Ends Well, Act 3, Scene 5).

The study of honesty continues to be relevant. From ethical scandals that expose dishonest politicians and business leaders, to misinformation and fake news in (social) media, to infidelity in relationships, acts of dishonesty are common. Understanding scientifically when people are honest and when they are not informs those fighting corruption, fraud, and public distrust. Such scientific efforts not only reveal why individuals choose to act truthfully or lie, but also identify conditions that promote honesty. Because trust is fundamental to societal stability, promoting honesty is essential to rebuilding trust in institutions and in each other. Policymakers, organizations, and individuals often use insights from this line of work to create transparent systems and promote ethical behavior.

Dishonesty can be driven by expressive or instrumental motivations (Dunning, 2015). When driven by expressive motivation, dishonesty fulfills immediate goals. Such behaviors are common and widespread. Indeed (for a review, see Hofmann & Grigoryan, 2023) examined morality in everyday life and found most immoral acts are associated with harm and injustice, meaning they may be motivated by the immediate goal of inflicting suffering on others. Similarly, dishonestly spreading misinformation is driven by the immediate goal of facilitating partisanship (Effron & Helgason, 2022). However, when dishonesty is driven by instrumental motivation, the goal is future-oriented and typically a means to another end—most often financial gain, for example, when considering embezzlement, corruption, and fraud.

Here, we review the current state of research on the science of (dis)honesty and suggest directions for future study. Our review focuses on (dis)honest behavior driven by instrumental motivations, such as financial gain, and is grounded in the behavioral ethics literature. In short, we argue that we know a lot about people's (dis)honesty in a very specific context, namely, when people make decisions that affect only themselves, without fear of being sanctioned for lying. We know much less about more realistic contexts in which lying may lead to sanctions, and in which people's lies may influence, or sometimes depend on, others. Our framework calls for more research on these less well-understood, but very relevant, contexts.

1. The state of the art: People are rather honest, but not always

Fifteen years of behavioral research on honesty suggests that although people tend to be quite honest, they are not always (Abeler et al., 2019; Stephens-Davidowitz, 2017). Two seminal studies (Fischbacher & Föllmi-Heusi, 2013; Mazar et al., 2008) have led the way in examining the motives behind people's propensity to lie. Using a variety of experimental tasks, both papers made an important finding: when individuals can lie to increase their own gains without risk of detection or punishment, many still refrain from lying to the maximum extent possible. Moreover, a significant fraction of individuals will resist lying even for substantial rewards, such as €150 for a single lie (Hilbig & Thielmann, 2017).

From a rational economic perspective, this behavior is puzzling. When no risk of detection or punishment exists, one would expect people to seize the opportunity to maximize their returns by lying. An important explanation for why people refrain from lying is their desire to maintain a positive self-image. In other words, even when people lie, they still want to perceive themselves as moral.

According to the psychological framework of self-maintenance (e.g., Mazar et al., 2008), dishonest behavior creates an internal conflict between two opposing goals. On the one hand, people highly value their morality (Greenwald, 1980; Sanitioso et al., 1990) and strive to maintain it as a central aspect of their self-image (Aquino & Reed, 2002; Rosenberg et al., 1995). On the other hand, when the opportunity to benefit from lying arises, resisting the temptation is difficult. Caught between a rock and a hard place, this internal tension—referred to as ethical dissonance—poses a serious threat to people's self-concept and moral identity (Barkan et al., 2015). The observation that people cheat, but only to a certain extent (e.g., Gneezy et al., 2018; Hochman et al., 2016; Mazar et al., 2008; Shalvi et al., 2011), reflects a solution whereby people benefit from lying while perceiving themselves as moral (Shalvi et al., 2015).

The extent to which people avoid lying depends largely on the context in which they face the temptation. When lying benefits only those who lie, people limit their lying (Abeler et al., 2019; Gerlach et al., 2019). However, a study by Leib et al. (2021) found that when lying requires collaboration—when two or more people must lie together—about one-fifth of their 2716 participants lied to the fullest extent possible, engaging in what has been termed "collaborative dishonesty" (Weisel & Shalvi, 2015).

Collaborating with others appears to provide a justification for lying, especially when doing so benefits the group to which one belongs (Shalvi & de Dreu, 2014). Similarly, when lying benefits others, even when collaboration is not required, people lie more than they would if lying did not benefit others (Conrads et al., 2013; Klein et al., 2017; Wiltermuth, 2011).

Self-serving justifications, such as benefiting another person, shape how much people lie. Consider, for example, the die-under-cup task used by Shalvi et al. (2011) to study the role of self-serving justifications (see Fig. 1). Building on the task introduced by Fischbacher and Föllmi-Heusi (2013), participants were asked to privately roll a die under a paper cup to determine their pay. Shalvi et al. (2011) paid participants according to the number they reported rolling; reporting a "1" meant earning US$1, reporting a "2" meant earning US$2, and so on until reporting a "6" resulted in earning US$6. If participants are honest, they should report the number they observed. With a large enough sample and a fair dice, the distribution of reported outcomes, assuming honesty, should be flat. A similar proportion of participants should report 1s, 2s, 3s, 4s, 5s, and 6s. If all participants are dishonestly trying to maximize their payoff, they should all report "6."

A key variation in the study was the number of dice rolls participants were able to observe before reporting. When participants rolled only once, their lying was minimal. However, when they were allowed to roll three times and report only the result of the first roll, lying increased. The logic is that with multiple rolls, participants could justify lying by reporting a higher number that appeared on one of the rolls that does not count for pay. In this scenario,

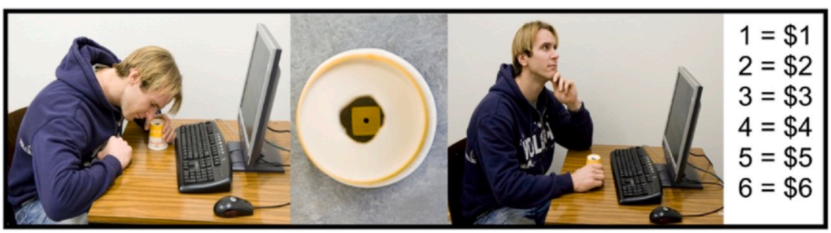

Fig. 1 Die-under-cup task. Participants roll a die under a cup, report the outcome, and receive payment according to the number they report. Because participants are the only ones who see the rolls, they can cheat to increase payoffs. *Reprinted from Shalvi, S., Dana, J., Handgraaf, M. J. J., & De Dreu, C. K. (2011). "Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior." Organizational Behavior and Human Decision Processes, 115, p. 184. Copyright 2011 by Elsevier.*

reporting a counterfactual—an observed higher number that did not count—seemed to provide participants with a moral buffer that made dishonesty feel more justified than simply making up a number. Relatedly, as the baseline probability of winning increases, people are more likely to lie and report winning (Schild et al., 2020).

Aggregate evidence suggests self-serving justifications significantly influence the extent to which people lie. A meta-analysis by Abeler et al. (2019) examined the die-rolling task described above, where participants were either instructed to roll the die once before reporting the result or were allowed to roll multiple times but only report the result of the first roll. Although this variation was not randomly assigned as in Shalvi et al. (2011), it still provided insight into how observing additional die rolls affects honesty. Meta-analytic results from a large sample ($N = 44{,}390$) suggest multiple rolls do indeed lead to higher, more dishonest reports. Participants who were asked to roll only once were significantly more honest than those who were allowed to roll multiple times.

Further support for the positive influence of self-serving justifications on lying comes from cross-cultural research ($N = 2568$ participants), which shows that in societies where rule-breaking is more common, participants' reports on the dice-rolling task (in which lying is assessed at the aggregate level) closely resembled the distribution that would be expected if they had reported the highest of the dice-roll outcomes that they observed, rather than the outcome of the first dice roll that they were supposed to report (Schulz & Gächter, 2016). This finding suggests that in cultures that are more tolerant of dishonesty, people are more likely to use self-serving justifications.

However, results from online adaptations of the dice-rolling task have been mixed. Some studies, such as Baro, Verschuere, and Nieper (working paper), found multiple rolls led to more dishonesty than a single roll, whereas others, such as Ay et al. (2024), did not observe this effect. These inconsistencies suggest potential differences between laboratory and online settings may influence the level of dishonesty and warrant further investigation to understand how context affects lying behavior (see Lilleholt et al., 2020).

Although the above summary of the current state of the art is not exhaustive, it represents the generally accepted way of conceptualizing and studying (dis)honesty. As we point out in the remainder of the manuscript, over the past 15 years, scholars have mostly (but not exclusively) focused on settings in which individuals can lie or be honest without being monitored and/or punished for lying. Our review suggests many interesting insights have emerged from such work and that several recent lines of work suggest expanding this scope can lead to a better understanding of when and why people lie.

2. Identifying gaps in knowledge: A new framework

A major gap in our understanding of honesty lies in the extensive focus on studying behavior in a "laboratory context"—specifically, a context in which individuals act in isolation and their lies cannot be detected. Much of the existing behavioral evidence on (dis)honesty comes from such a laboratory setting. A key advantage of such a setting is that it allows researchers to examine how image concerns shape dishonesty. The premise is straightforward: if people refrain from lying even when they know their lies will not be detected or punished, they are likely concerned with maintaining a positive self-image. Accordingly, limiting one's dishonesty likely reflects a desire to maintain an honest self-concept (Hochman et al., 2021).

Although these findings are intriguing, social psychology, experimental economics, and related disciplines aim to model real-world behavior through laboratory experiments. Outside the lab, the probability of being caught in a lie is rarely zero (although it can sometimes be very low), and people do not know what that probability is (Thielmann & Hilbig, 2018). Moreover, dishonesty in everyday life is often relational. People's lies often involve others, either through collaboration or by relying on others' tolerance of deceptive behavior.

To address these shortcomings, we introduce a framework that includes two dimensions: enforcement, or the likelihood that lying will be detected and punished (ranging from low to high), and the context in which decisions are made (ranging from solitary to relational), as shown in Fig. 2. The figure shows most research focuses on the upper-left quadrant—where enforcement is low and individuals make decisions in isolation—whereas much less is known about the other quadrants. Our proposed framework encourages the field to explore more realistic settings where enforcement mechanisms are in place (and punishment is possible) and where decisions may involve larger groups. We demonstrate the utility of this framework by reviewing relevant research that provides (preliminary) support for the key ideas we propose in each section, and by identifying open research questions in psychology and related fields.

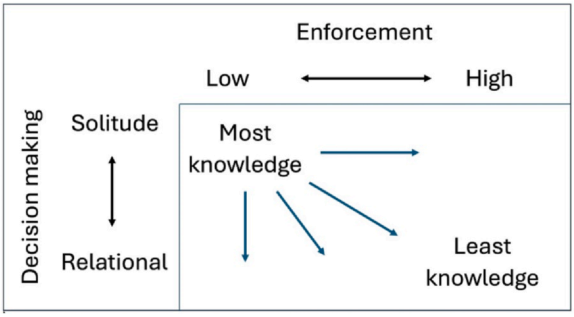

Fig. 2 Novel conceptual framework: (dis)honesty as a function of level of enforcement and decision-making context.

The remainder of the paper is organized as follows: Section 2 introduces a framework that highlights both well-researched areas and gaps in the literature; Section 3 reviews findings on (dis)honesty in five key areas and shows how the framework can bridge existing knowledge gaps. These areas include the structure of relationships between individuals (Section 3.1: interdependence, intragroup processes), relationship aspects (Section 3.2: social norms, trust, gossip); individual differences (Section 3.3: personality, beliefs, justifications, dynamic personality theories, care for honesty vs. prosociality, gender), cognitive processes (Section 3.4: ethical dissonance, attention, perspective-taking, automatic vs. controlled behavior, embodied cognition), and situational factors (Section 3.5: honesty pledges, artificial intelligence). Finally, Section 4 provides concluding remarks.

2.1 Enforcement and compliance

Enforcement mechanisms—acts that compel compliance with a rule—can reduce dishonesty. Enforcement is typically viewed as the product of the probability of detection multiplied by the severity of the punishment for dishonesty (Becker, 1968). Although empirical crime studies have found consistent evidence that the probability of punishment is more important than its severity (Chalfin & McCrary, 2017; Doob & Webster, 2003; Nagin, 2013), most experimental studies have found no effect for the probability of detection (e.g., Mazar et al., 2008) or have found stronger effects for the severity of punishment than for the probability of detection (e.g., Anderson & Stafford, 2003; Thielmann & Hilbig, 2018). One possible explanation for this discrepancy is the gap between these experimental studies and real-world settings. Most research on dishonesty focuses on tasks in which people have limited opportunities to lie, meaning studies of sustained deterrence are relatively rare (Gerlach & Teodorescu, 2022). In real-world situations, however, decisions are often repeated, making an understanding of sustained deterrence crucial for predicting dishonest behavior outside of the laboratory.

Focusing on sustained deterrence, Teodorescu et al. (2021) manipulated both the probability of detection and the amount of punishment, while holding the expected punishment constant. The authors created environments in which participants repeatedly encountered temptations to increase their payoffs by providing dishonest answers. Initially, participants provided answers without external enforcement, allowing a baseline level of honesty to be established. Later, an enforcement mechanism was introduced: participants' answers were checked with a certain probability, and dishonest reports resulted in fines. The study compared two enforcement policies: infrequent large fines versus frequent small fines. Both enforcement strategies increased honesty, but frequent small fines were far more effective, especially among the most dishonest participants in the no-enforcement phase.

These results suggest the frequency of enforcement is a stronger predictor of honesty than the severity of punishment. This pattern of results is consistent with empirical studies of crime, which show crime rates are more sensitive to the frequency of rule enforcement than to the severity of punishment (Chalfin & McCrary, 2017; Doob & Webster, 2003; Nagin, 2013). Thus, studying environments with repeated temptations to lie provides a more ecologically valid setting for understanding how external enforcement shapes honesty.

In real-world contexts, the frequency of enforcement is not constant and may vary across time and place. This scenario can result from adjustments in enforcement resources, such as increasing or decreasing surveillance or shifting focus to "crime hot spots" (Weisburd et al., 2023). As summarized in Table 1, a critical question for future research is the extent to which dishonest behavior adapts to changes in enforcement frequency. The findings of Teodorescu et al. (2021) suggest individuals may be highly responsive to very low detection rates, but less adaptive to changes when enforcement is moderate or high. Interestingly, research on tax evasion has shown the experience of being audited and punished shapes people's decisions to cheat in the future. For example, Kastlunger et al. (2009) found compliance decreases after being audited. This behavior is referred to as the "bomb crater effect," according to which people tend to underestimate the likelihood of being punished shortly after enforcement (Mittone, 2006; see also Plonsky et al., 2015).

Table 1 Open research questions.

Section	Subsection	Open research questions
2.1 Enforcement and compliance	2.1.1. The Effectiveness of Enforcement in Promoting Honesty	1. How adaptive is dishonest behavior to changes in enforcement frequency? ·Depending on the range of enforcement frequencies (e.g., low vs. high) ·Depending on the speed of the change (e.g., gradual vs. abrupt)? 2. How does being caught engaging in dishonest behavior influence the perceived likelihood of being caught again in the near future? 3. When detection of dishonest behavior does not necessarily result in punishment, which factor is more important: the likelihood of detection or the likelihood of punishment? 4. Which types of punishment minimize potential negative effects of increased deviance? 5. Do law enforcers and educators reduce the frequency of enforcement when faced with severe penalties? And if so, do potential violators exploit this leniency?
3.1 Relationship structure	3.1.1 Interdependence	6. How does deception driven by self-serving motivations compete with motivations for equity? 7. How do the characteristics of the "victim" (e.g., their identifiability or gender) influence deception?
	3.1.2 Intragroup Processes in Dishonesty: From Cooperation to Culture	8. How a culture of cheating manifests itself in different ways at work, in sports, at school, in arts, etc. 9. How routinized cheating leads to reduced perception of guilt. 10. How highly competitive environments—e.g., in the industry or science—may lead to collective moral disengagement (Bandura, 2002), e.g., under the form of beliefs that "the end justifies the means."

	3.1.3 Intergroup Processes in Dishonesty	11. How groups' gender composition influences dishonesty. 12. Interaction between types of dishonesty and group membership(s). 13. Whether punishment is effective in combating dishonesty in a highly relational context.
3.2 The social glue binding relationships	3.2.1 How social norms affect and are affected by (dis) honesty	14. Dynamic relationship between changes in social norms and changes in dishonesty. 15. How "sticky" norms are in the context of dishonesty. 16. How to leverage norm-nudging to reduce dishonesty.
	3.2.2. Trust	17. What is the optimal level of trust between people and institutions and how can such an optimum be conceived? 18. How can we define the relationship between polarization, compliance, and trust? 19. What are the proximate and ultimate factors contributing to trust? 20. Is trust a universal human concept, similar across cultures, potentially even shared with our closest living relatives—the great apes and their common ancestor—or is it not? 21. What is the role of emotion and its various forms of expression in trust formation?
	3.2.3 Gossip	22. How does (potentially) inaccurate gossip affect honesty? 23. How do different patterns of interdependence in the gossip triad affect honesty? 24. Can gossip be used to identify partners for collaborative cheating? 25. What are the effects of becoming the target of gossip (rather than merely experiencing the threat of gossip) on honesty?

(continued)

Table 1 Open research questions. (*cont'd*)

Section	Subsection	Open research questions
3.3 Individual differences	3.3.1 Personality, beliefs, and justifications	26. Do unethical individuals adopt justifying beliefs, or do beliefs gradually incline people toward unethical behavior (or both)? 27. Which social-context factors shape justifying beliefs, especially those referring to others' ethicality and the probability/effectiveness of sanctions?
	3.3.2 Honesty and Personality Dynamics	28. Which methods are best suited to capture meaningful honesty and dishonesty in daily life, and what are the pros and cons of different methods? 29. What are the personality mechanisms underlying moral traits such as honesty? 30. What kinds of real-life goals are generally facilitated by honesty, by dishonesty, or potentially by either one (depending on the situation)? 31. To what degree do people differ in terms of valuing, believing in their ability to attain or avoid, or perceiving the availability of particular (dis)honesty-relevant goals? 32. To what degree can person-level differences in (dis)honest behavior be explained by goal-related dynamics?
	3.3.3. Individual Differences in Trading Off Honesty Against Pro-sociality	33. How compliance is situated between honesty and pro-sociality?
	3.3.4 Gender differences in lying	34. Does the possibility of detection have a greater impact on increasing honesty in women than men? 35. Do group interactions increase the likelihood of women lying due to exposure to men's arguments in favor of dishonesty?

3.4 Cognitive processes	3.4.1 Ethical Dissonance	36. What is the trade-off between the internal costs of ethical dissonance and the benefits of external enforcement? 37. Do external incentives reduce intrinsic motivation in contexts involving dishonest behavior? 38. Can internal enforcement mechanisms (such as moral reminders or visibility cues) enhance the effectiveness of existing external enforcement policies? How can internal and external enforcement be combined to achieve maximum deterrence? 39. How does group involvement affect an individual's experience of ethical dissonance? a. Under what conditions do groups alleviate an individual's ethical dissonance, leading to disengagement? b. Under what conditions do groups increase an individual's ethical dissonance, thereby heightening perceived visibility and self-awareness? 40. How can we measure collective ethical dissonance, and how can this dissonance function as a moral gatekeeper at the group level?
	3.4.2 Attention	41. Are ethical blind spots more likely to emerge in corrupt environments? 42. Can observing others cheat make people ethically blind? 43. Do people show blind spots when pressured to do wrong without personally benefiting from dishonesty? 44. Are blind spots more common in corrupt collaborative settings when incentives are aligned? 45. Do people overlook the dishonesty of others when they believe it is justified? 46. Does enforcing honesty reduce people's tendency to overlook justifiable and unjustifiable dishonesty of others?

(continued)

Table 1 Open research questions. (*cont'd*)

Section	Subsection	Open research questions
	3.4.3. Perspective	47. What constitutes a lie, and what is perceived as a lie? (When do people view it as "not lying," a "small lie," or something that is acceptable within social norms?). 48. What is the underlying cognitive process that assists in maintaining one's self-image in case of cheating?
	3.4.4. Honesty as an automatic versus controlled behavior	49. Is there a simpler alternative to resting-state connectivity to reliably assess moral default and explain heterogeneity in self-report, behavioral, and neural responding to (dis)honesty? 50. Do current findings, typically obtained in testing individuals in isolation, expand to real-time, dynamic two-way interaction? 51. And does the science on the intuitive nature of (dis)honesty allow development of new ways to tell lie from truth? 52. Do emerging insights on the intuitive nature of (dis)honesty allow development of interventions to promote greater honesty in society? 53. Does the complexity of lying affect the relationship between self-control and the act of lying?
	3.4.5. Embodied (dis)Honesty	54. Can changes in bodily self-consciousness be used to promote honesty/reduce dishonesty? 55. Can modulation of brain regions related to BSC be used to promote honesty/reduce dishonesty?

3.5 Situational factors	3.5.1. Honesty Pledges	56. To what extent can the effects of honesty pledges observed in online or laboratory settings be generalized to real-life situations, and what are the barriers and effectiveness to implementing honesty pledges in various real-world contexts, such as tax reporting, public service, or business ethics? 57. What specific elements of honesty pledges (e.g., wording, format, timing) are most effective in reducing dishonest behavior? 58. How do different situational factors (e.g., punishment severity, task repetition, public vs. private decision-making) influence the effectiveness of honesty pledges? 59. How can honesty pledges be integrated into existing legal and administrative systems to reduce the administrative burden and improve compliance? 60. Do the effects of honesty pledges persist over the long term, or do they diminish with time, and how frequently should honesty pledges be renewed or reinforced to maintain their effectiveness? 61. What are the psychological mechanisms that underlie the effectiveness of honesty pledges? 62. Can honesty pledges foster mutual trust between governments and citizens, managers and employees, or within communities, and what are the potential unintended consequences of honesty pledges, such as increased skepticism or resistance if perceived as coercive?

(continued)

Table 1 Open research questions. (*cont'd*)

Section	Subsection	Open research questions
	3.5.2 Artificial intelligence	63. How does AI, as a team member, influence groups' (dis)honesty, and how do different group compositions shape it? a. How do different proportions of AI agents influence a group of humans?b. Is the effect symmetric when comparing honest and dishonest AI agents? 64. How does AI, as a managing principal, form groups and influence group dynamics to promote honesty? 65. When and why are blame and responsibility for (dis)honesty shifted away from individuals toward AI agents? What are the consequences of such a shift? 66. How does the presence of AI detection tools impact individuals' ethical behavior and when are AI tools perceived as legitimate in such detection roles?

Furthermore, does the speed at which enforcement changes matter? Comparing responses to abrupt versus gradual changes in detection probability could shed light on how sustained deterrence evolves. One promising avenue for addressing these questions is studies of decision-making in dynamic environments (Hochman & Erev, 2013; Plonsky et al., 2015; Teodorescu & Erev, 2014), which suggest people adjust their behavior more quickly to abrupt than to gradual changes. Examining how changes in the frequency of enforcement affect dishonesty could provide valuable insights for policymakers and deepen our understanding of the processes underlying dishonest behavior.

Finally, work on enforcement from a legal perspective highlights that in real life, and especially in the context of serious crimes, detection does not necessarily imply punishment (see Nagin, 2013). In laboratory settings, these factors may be more confounded than in the real world, where the detection of a lie and the determination and execution of punishment are not givens. In addition, criminological research has found a difference between objective and subjective deterrence, as people may over- or underestimate the likelihood of being caught and punished (Decker et al., 1993; van Rooij, 2016). Such research has examined how deviant decisions update information after word of punishment (Apel, 2013), something laboratory research on honesty and deterrence could also do.

In this context, we should remember deterrence is one of many potential effects of punishment on behavior. We know punishment can have both positive and negative effects on illegal or deviant behavior. Positive effects that may reduce deviance include specific and general deterrence (Nagin, 2013; Nagin et al., 2009), assurance of compliance (Thornton et al., 2005), and signaling what the norm is van Rooi et al. Negative effects that increase deviance include crowding out prosocial norms (Gneezy & Rustichini, 2000), activating antisocial norms (Cialdini et al., 2006), stigmatization (Feingold, 2021), and undermining legitimacy in the case of procedurally unfair punishment (Tyler, 2006; Walters & Bolger, 2019). Future research on honesty that seeks to understand the effects of enforcement should consider these broader effects.

2.1.1 From effective enforcement to compliance

An ongoing challenge for policymakers seeking to improve compliance through effective enforcement is the trade-off between the likelihood of detection and the severity of punishment. This dilemma arises from the potentially counterproductive nature of harsh penalties. If enforcers perceive harsh penalties as unjust, their willingness to consistently enforce rules may decrease (Feess et al., 2018), leading to reduced enforcement activity and, consequently, a lower probability of detection. In this case, several critical questions arise: Do enforcers actually reduce the frequency of enforcement when faced with severe penalties? If so, do potential violators take advantage of this leniency and increase their dishonest behavior when they expect severe punishment but perceive lax enforcement? The interaction between those who enforce the rules and those who are monitored remains an important area for further research.

Policymakers should prioritize building trust, enhancing legitimacy, and fostering a regulatory environment that encourages voluntary compliance while maintaining effective enforcement mechanisms for those less inclined to comply. Although honesty is often viewed as a personal virtue, compliance with laws and regulations is a social behavior influenced by individual characteristics, social factors, and institutional contexts (Feldman, 2018). By understanding the complex interplay between personal honesty, institutional trust, and compliance motivations, authorities can develop more effective strategies to promote law-abiding behavior among diverse populations (Becker, 1968; Frey & Jegen, 2001).

Adopting a responsive regulatory approach, as suggested by Braithwaite (2002), involves tailoring regulatory responses to the motivations and behaviors of different individuals or groups, which can increase the likelihood of policy success. For example, authorities might begin with a cooperative approach and escalate enforcement only when faced with noncompliance. This strategy helps build trust with those who are inclined to comply voluntarily, while maintaining a deterrent effect on potential violators.

In summary, Fig. 2 illustrates the knowledge required to understand how enforcement shapes honest behavior. As the figure further illustrates, knowledge of honesty as a relational process is also required, which we discuss in the next section.

2.2 Honesty as a relational process

Honesty is inherently relational, involving a dynamic interaction between communicators—individuals, groups, or organizations—and the recipients of the communication (Cooper, Cohen, Huppert, Levine, & Fleeson, 2023). What a communicator chooses to share and how they share it significantly influences whether their audience forms accurate beliefs. However, the responsibility for honesty does not rest solely with the communicator, as they are only one side of the interacting parties. Effective honesty requires not only a communicator (the actor) but also a receiver whose psychology, behavior, knowledge, skills, and abilities also affect the likelihood that all parties involved in the communication will reach a mutual understanding of the truth (Cooper, Cohen, Huppert, Levine, & Fleeson, 2023; Ren et al., 2022).

In some cases, such as self-deception, the actor and the receiver may be the same person, and the audience of the communication is oneself (Schwardmann & Van der Weele, 2019; Von Hippel & Trivers, 2011). This situation is unique. More commonly, the audience includes external parties such as friends, family members, romantic partners (DePaulo & Kashy, 1998; Levine & Cohen, 2018; Reinhardt & Reinhard, 2023), colleagues, clients, and others with whom one interacts professionally (Bourdage et al., 2018; Keep, 2009; Shell, 2021; Yagil & Medler-Liraz, 2014), shareholders (Yuthas et al., 2002), social media followers, and the general public (Pennycock et al., 2020).

Viewing honesty as a relational process with its own relational motives and consequences underscores why the dominant methods of studying honesty over the past 15 years may be limited. Stylized experiments that focus on binary decisions, such as lying to an experimenter or telling the truth for a small monetary reward (Abeler et al., 2019; Gerlach et al., 2019), fail to capture the complex social dynamics that influence honesty in real-world interactions. These controlled experiments, which emphasize simplicity and clean experimental designs, often ignore the intricate social and relational contexts in which honesty and dishonesty naturally occur. As a result, these studies treat honesty as an isolated individual choice, divorced from the potential for detection and devoid of relational or reputational consequences. This approach is useful for understanding how, when, and why people lie in solitude, but it conspicuously sidesteps large, important questions related to truth-seeking, authentic communication, and fostering mutual understanding of truth in others (as discussed by Cooper et al., 2023 see Fig. 3).

The use of the sender-receiver game introduced by Gneezy (2005) may be a first step in this direction. In this stylized game, the sender is given private information and must report it to the receiver, either truthfully or not. The receiver then makes a decision. The payoffs for both the sender and the receiver depend on the receiver's decision. In a common variant, the receiver must guess the true state of the world, with payoffs for both players depending on whether this guess is correct (Erat & Gneezy, 2012). This game has proven useful for studying some aspects of relational lying, particularly sophisticated deception, in which the sender tells the truth with

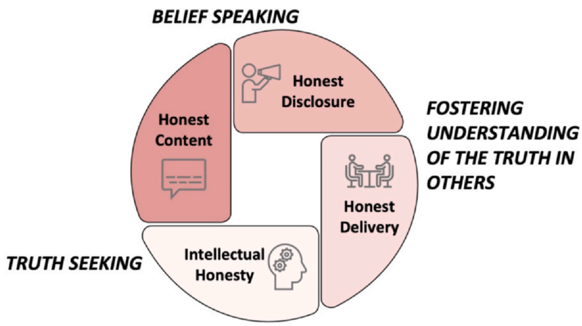

Fig. 3 A Multidimensional framework of honest behavior. **Notes**: "Honest content" refers to the accuracy of the content a communicator shares. "Honest disclosure" refers to the amount and depth of disclosure by the communicator. "Honest delivery" refers to the manner in which the communicator shares or discloses information. "Intellectual honesty" refers to the process by which a communicator develops, validates, and updates their beliefs. *Reprinted from Cooper, B., Cohen, T. R., Huppert, E., Levine, E. E., & Fleeson, W. (2023). Honest behavior: Truth-seeking, belief-speaking, and fostering understanding of the truth in others. Academy of Management Annals, 17(2), 655–683, printed with permission from the Academy of Management.*

the intent to deceive and expects the receiver not to believe him (Cohen, Gunia, Kim-Jun, & Murnighan, 2009; Blazquiz-Pulido et al., 2024; Sutter, 2009). However, numerous other aspects of relational lying remain unexplored and could be further explored using this or other games as basic tools.

To advance our understanding of honesty in everyday life, adopting research methods that reflect the relational nature of honesty is critical. Adopting such methods means examining how people navigate truthfulness in ongoing relationships, considering factors such as relationship history, social norms, and the potential long-term consequences of honesty or dishonesty. It also means exploring how honesty is maintained or compromised in different relational contexts, such as families, workplaces, or broader social networks, and how different stakeholders interpret and respond to honest or dishonest communication.

By recognizing honesty as a dynamic, contextual process rather than a static, isolated act, we can gain deeper insights into the factors that promote or inhibit truthful communication. This broader perspective not only enriches our theoretical understanding, but also has practical implications for designing interventions and policies to promote honesty and trust in various social and organizational settings.

A systematic review of 169 empirical studies of honesty and dishonesty in management, organizational behavior, applied psychology, and business ethics conducted between 2000 and 2021 found nearly three-quarters of the articles focused primarily on "the accuracy of the content a communicator shares" (i.e., honest content) (Cooper et al., 2023). Far fewer studies examined other critical dimensions of honesty, such as "the amount and depth of disclosure by the communicator" (i.e., honest disclosure), "the manner in which the communicator shares or discloses information" (i.e., honest delivery), and "the process by which a communicator develops, validates, or updates his or her beliefs" (i.e., intellectual honesty). If a similar review were conducted in the field of social-personality psychology, we would likely observe the same trend: an overwhelming focus on the antecedents of accurate versus inaccurate reporting, with much less attention paid to the nuanced ways communicators can foster false beliefs in their relational partners through both verbal and nonverbal strategies beyond outright lying.

These overlooked strategies include dodging, deflecting, hedging, remaining silent, omitting key information, avoiding certain topics, and neglecting to fact-check. Such behaviors—strategic or inadvertent—can contribute significantly to misunderstandings and misinformation in interpersonal relationships (Bitterly & Schweitzer, 2019; Levine, Roberts, & Cohen, 2020). The role of the recipient of communication is equally important: when recipients fail to ask questions, raise concerns, or exercise due diligence, they may inadvertently contribute to the perpetuation of false beliefs and unchallenged dishonesty.

A growing body of research suggests that whereas communicators often view these indirect tactics as less unethical than outright lying, recipients do not (see also Levine et al., 2018; Park & Klein, 2024; Levine & Duncan, 2020). Recipients tend to be highly sensitive to whether a communicator is intentionally fostering misunderstanding, regardless of the specific tactic employed. For example, communicators may view prosocial lies of omission—such as withholding inconvenient truths—as more acceptable than prosocial lies of commission, in which false information is actively provided. Recipients, however, often take the opposite view and perceive omissions as equally, if not more, deceptive.

Similarly, when communicators choose to "ghost" a relationship partner—effectively ending communication without explanation—they may believe they are sparing the other person's feelings by avoiding a difficult conversation. However, those who are ghosted typically do not interpret this action as kind or considerate, and instead may experience increased confusion and distress. These dynamics underscore the important role that relationships and different perspectives play in shaping judgments of (dis)honesty.

In sum, the dominant focus of research on honest content overlooks the complex relational processes through which honesty and dishonesty manifest. By broadening the scope of inquiry to include the subtleties of honest disclosure, delivery, and intellectual honesty, scholars can gain a more comprehensive understanding of how (dis)honesty operates in interpersonal and organizational contexts. This holistic perspective is essential for developing effective interventions and policies to promote honest and transparent communication.

2.2.1 Motives for (dis)honesty

Understanding honesty as a relational process illuminates the diverse motives that drive both honest and dishonest behavior. In controlled laboratory settings, participants often lie for self-interested reasons or to achieve mutual benefits, as seen in corrupt collaboration paradigms (Weisel & Shalvi, 2015, pp. 20–22) and research on the interindividual-intergroup discontinuity effect (Cohen, Gunia, Kim-Jun, & Murnighan, 2009; Wildschut & Insko, 2007). In particular, group decisions tend to be influenced by motivations such as greed and strategic considerations aimed at maximizing collective outcomes. In addition, fear and distrust often lead to deceptive behavior, driven by the belief that one will be met with similar distrust from others (Cohen et al., 2009; Sutter, 2009; Wildschut & Insko, 2007) or simply because cynical beliefs provide a particularly powerful justification (Hilbig et al., 2022).

Outside the laboratory, however, the motives for lying extend far beyond self-interest and fear. Individuals often engage in dishonest behavior due to complex relational factors, including spite (Erat & Gneezy, 2012), competitiveness (Hilbig et al., 2024), loyalty (Hildreth & Anderson, 2018; Thielmann et al., 2021), prosocial intentions (Klein et al., 2017; Levine & Schweitzer, 2014, 2015; Wiltermuth, 2011), harm avoidance (Levine, 2022), and paternalism (Lupoli, Levine, & Greenberg, 2018). Although these motivations can be studied in experimental settings, fully understanding why people choose to act on them requires acknowledging the role of relationships in shaping communication.

Lies motivated by relational concerns often lead to more nuanced outcomes than those motivated by self-interest. For example, lies that are perceived as unambiguously prosocial—helping the recipient without causing harm—are generally considered ethical (Levine & Schweitzer, 2014) and may even increase trust (Levine & Schweitzer, 2015). By contrast, self-serving lies are typically viewed as unethical and detrimental to trust. However, not all prosocially motivated lies are perceived positively. When a communicator lies to benefit the recipient but lacks sufficient understanding of the recipient's preferences, the lie is perceived as paternalistic rather than prosocial. For example, offering false hope to a patient who values optimism over realism might be seen as a clearly prosocial act. However, offering the same false hope to a patient who has not expressed such a preference will be seen as paternalistic. Paternalistic lies, like self-interested ones, are often judged to be unethical and are more likely to be punished (Lupoli et al., 2018). This distinction highlights the importance of a communicator's awareness of the recipient's preferences in shaping judgments of dishonesty.

Context also plays a critical role in judgments of dishonesty. Recent research suggests people are more likely to condone dishonesty that prevents unnecessary harm. Judgments of unnecessary harm are based on two key inferences: the extent to which honesty would cause immediate emotional harm to the recipient, and whether it holds potential for long-term learning and growth (i.e., its instrumental value). When honesty is perceived as causing emotional harm without offering instrumental value, people are more likely to view it as causing unnecessary harm, thereby justifying dishonesty as the more ethical choice. Certain contexts are particularly conducive to judgments of unnecessary harm; for example, when the recipient is unable to understand, respond to, or learn from painful information, providing it may be seen as unnecessarily harmful. Consider the scenario of giving someone negative performance feedback: although critical feedback can cause emotional distress, it also provides an opportunity for the recipient to learn and improve in the future (Abi-Esber et al., 2022; Fulham et al., 2022). In cases where the recipient cannot benefit from such feedback, false praise may be seen as the more ethical response (Levine, 2022).

2.2.2 The case of corrupt collaboration

An important example of the relational aspect of dishonesty is corrupt collaboration, or the use of joint acts of dishonesty to achieve personal gain (Leib et al., 2021; Weisel & Shalvi, 2015). The main finding in this line of work is that collaborative settings, in which agents work together toward common goals and can benefit from each other's dishonest behavior, are particularly prone to excessive levels of dishonesty. For many people, collaboration as a noble cause seems to offset the moral costs of dishonest behavior, allowing them to be dishonest while maintaining a reasonable moral self-image. In this view, cooperation and honesty can be seen as moral currencies that can be traded for one another (Weisel & Shalvi, 2022).

The notion of tradable moral currencies fits well with the relational aspects of (dis)honesty highlighted in the current framework. When multiple people are involved in acts of (dis)honesty, determining the total moral "value" of such acts naturally requires taking into account the welfare of all the agents involved. In this view, even acts that are at face value immoral, for example, because they involve lying or harming others, may ultimately be judged as morally acceptable by both decision-makers and others, depending on the outcome of the moral calculus.

An understudied, relational aspect of corrupt cooperation is the role of collaborators' group affiliations. What are the dynamics of corrupt cooperation among members of different groups? How do group characteristics, whether actual or perceived, and intergroup relations affect corrupt cooperation? Dorrough et al. (2023) examined similar questions in the context of bribery, and seeing what patterns emerge with respect to corrupt collaboration would be interesting.

One of Weisel and Shalvi's (2015) findings is that dishonesty is most prevalent when both members of a dyad stand to gain equally from collaborative dishonesty, as opposed to situations where dishonesty is more profitable for one dyad member than the other. The notion that symmetry between dyad (or group) members can encourage dishonesty can be applied and tested in the current framework by examining symmetric and asymmetric enforcement mechanisms (see Fig. 2) in the context of corrupt collaboration. Both the probability of detection and the severity of punishment can vary across potential collaborators. A simple hypothesis, based on the results of Weisel and Shalvi (2015), is that dishonesty will be more prevalent when enforcement is identical for all participating agents than when it is asymmetric. Testing the effect of asymmetric enforcement on corrupt collaboration is not just an intellectual exercise; because collaborating agents may be of different ranks, members of different organizations, have different levels of access to legal recourse, or be of different nationalities, considering asymmetric enforcement mechanisms can add a layer of external validity to future work.

A related avenue for future work can build on the work of Teodorescu et al. (2021), who systematically varied the frequency and severity of punishments in an individual (dis)honesty task and found frequent small fines were a stronger deterrent to dishonesty than infrequent large fines. Testing whether the corrupting power of cooperation renders such differences between enforcement mechanisms would be interesting.

3. Using our novel framework to reevaluate the science of honesty

As highlighted in Fig. 2, the current review focuses on the interplay between enforcement and the relational aspects of (dis)honesty. In Section 3, we assess how such a perspective can shed new light on five major strands of knowledge in the study of (dis)honesty. We do so by reviewing the state of the art in each of these strands of knowledge and highlighting open questions. These strands of knowledge about (dis)honesty include relationship structure, relationships features, individual differences, cognitive processes, and situational factors.

3.1 Relationship structure

3.1.1 Interdependence

Interpersonal interdependence refers to "the ways in which two individuals influence each other's outcomes in the course of their interaction". In the context of (dis)honesty, interdependence occurs when two individuals influence each other's decisions and outcomes in situations involving potential deception or truthfulness. According to interdependence theory (Kelley et al., 2003; van Lange & Balliet, 2015), interdependent situations are characterized by six key characteristics: (a) dependence: the extent to which one's outcomes are dependent on the actions of one's interaction partner; (b) power: the extent to which one has control over one's own and one's partner's outcomes; (c) conflict: The extent to which the best outcome for an individual results in a worse outcome for their partner; (d) coordination: how one individual's behavior affects the partner's actions and, consequently, one's own outcomes; (e) future interdependence: the impact of current behaviors on both partners' future interactions and outcomes; and (f) information certainty: the degree to which both partners are aware of each other's preferences and the consequences of their actions.

The two types of dishonesty we consider—individual and relational—differ significantly along these dimensions. Individual dishonesty occurs when a person lies to benefit themself at the expense of others, as in the classic die-rolling task, where participants may misreport the outcome of a private die roll for personal gain (Gerlach et al., 2019; Shalvi et al., 2011). This form of dishonesty is characterized by low interdependence and coordination, because the actions of others do not affect one's own outcomes. It also exhibits high levels of power and conflict, because the individual's financial gain is solely under their control and comes at the expense of others.

By contrast, relational dishonesty, such as collaborative dishonesty, involves two or more individuals jointly engaging in dishonest acts, such as in the dyadic die-rolling paradigm (Leib et al., 2021; Weisel & Shalvi, 2015) or bribery behavior (e.g., Dorrough et al., 2023). This form of dishonesty demonstrates a high degree of interdependence, where the actions of the other influence the outcomes of each participant. It is characterized by high interdependence (the monetary reward in collaborative tasks depends on both participants' reported outcomes, making each person's payoff dependent on the other's behavior), high coordination (success requires the coordination of dishonest actions; see Dorrough et al., 2023; Ścigała et al., 2019; Zickfeld et al., 2023), low conflict (the best outcome for an individual is directly linked to the best outcome for his or her partner, resulting in shared interests and reduced conflict), and shared power (control over the final outcome is shared between both partners, as each influences the other's potential rewards).

Although evidence shows interdependence influences dishonest behavior (e.g., Weisel & Shalvi, 2015; but see Zickfeld et al., 2023), research is still in its early stages, particularly regarding the role of future interdependence and information certainty. For example, how collaborative dishonesty evolves over longer periods of time, such as months or years, remains an open question. Although experimental settings can simulate collaborative dishonesty, such behavior in real-world contexts may be driven by factors such as mutual trust, partner vulnerability, or the threat of retaliation—elements that typically develop over time.

In addition, how physical proximity versus online interaction affects the propensity to engage in collaborative dishonesty is unclear. Greater trust and rapport may develop in face-to-face interactions, potentially increasing the likelihood of dishonest collaboration. Conversely, face-to-face interactions may also increase shame and reputation concerns, limiting the extent of dishonest collaboration. Future research should also examine how individuals' preferences, along with those of their interaction partner, shape the development of collaborative dishonesty over time, and whether these preferences change based on prior experiences with dishonesty.

Finally, examining how factors such as group membership or stereotyping affect individuals' levels of honesty over time would be valuable. Understanding the dynamics of collaborative dishonesty in long-term interactions could inform the development of interventions aimed at reducing dishonest behavior in such contexts. By exploring these dimensions of interdependence, future studies can provide a more nuanced understanding of how and why dishonest behavior emerges and persists in both individual and collaborative settings.

3.1.2 Intragroup processes in dishonesty: From cooperation to culture

Focusing on the intra-group processes that shape collaborative dishonesty, Zanetti and Butera (2023) conducted a series of interviews that revealed individuals often recall acts of collective cheating as complex, highly organized group activities. Such cheating behaviors are often premeditated, with accomplices selected from close associates, making cheating, in some cases, a routine practice. Group members carefully assess potential risks and establish concealment strategies, including codes of silence, to protect their misdeeds. Notably, these cheating events, particularly in academic settings, are remembered not only as acts of dishonesty but also as opportunities for cooperation, loyalty, and group cohesion. This shared experience contributes to the formation of a collective identity, providing a psychological basis for justifying their dishonest actions as part of group solidarity.

The role of shared identity in facilitating dishonesty is further supported by experimental research by Rullo et al. (2024). Their studies showed participants were more likely to cheat when their group members shared a common identity, such as a national identity, than when they did not. In addition, participants were more likely to cover up the dishonest actions of a group member who shared their identity than those who did not. This finding suggests working with individuals who are perceived as similar encourages not only the act of cheating, but also the enforcement of a collective code of silence to protect fellow cheaters.

The tendency to prioritize and protect similar others in dishonest contexts parallels a broader adherence to certain value systems, as evidenced by Pulfrey et al. (2018) research on benevolence values. These values, which emphasize the welfare of close in-group members, predict positive attitudes toward collective cheating and even active participation in such behaviors, particularly in competitive settings. Conversely, Pulfrey and Butera (2013) found individual cheating behavior is positively associated with self-enhancement values, such as power and achievement, which are prevalent in neoliberal societies. On the other hand, adherence to self-transcendence values, such as universalism, which emphasizes the well-being of all humans and the natural world, negatively predicts individual cheating behavior (Pulfrey & Butera, 2016).

Cultural factors, both at the local and national level, significantly influence the prevalence of cheating. To better understand and measure these cultural influences, Zanetti and Butera (2025) developed the Culture of Collective Cheating Scale. This instrument is designed to capture the extent to which a collective culture of cheating exists within a particular institution or organization. It provides valuable insights into the tacit approval of dishonest behavior within specific cultural contexts and serves as a practical tool for educators, managers, and consultants. Stakeholders can use the scale to diagnose their organization's cheating culture and develop targeted interventions to reduce dishonesty.

An important question is: when do cheating and dishonesty become part of an organizational culture? Such an organizational culture consists of, and should be analyzed as, an interplay of explicit and implicit structures, values, and practices that are shared and transmitted among organizational members (see Schein, 2010; Tobsch et al., 2024). The fact that cheating exists collectively in an institution or organization does not mean it is part of its culture. For cheating to become part of an organization's culture, the organization would have to have accepted cheating in its values and transmitted as a norm to new members of the organization. Distinguishing organizational characteristics or practices from organizational culture is important because culture refers to something that is deeply rooted, not easily visible or known, even by organizational members, and very difficult to change (see Tobsch et al., 2024).

The body of research on intragroup processes underscores the importance of considering not only individual motives, but also the broader social and cultural contexts that facilitate and perpetuate dishonest behavior. Understanding the intricate web of group dynamics, shared identities, and cultural values that underlie collective cheating is critical to developing effective strategies for addressing dishonesty in various institutional settings. This research also draws attention to the understudied area of collective processes in cheating (see Fig. 2). In terms of enforcement, the available evidence points to two possible outcomes. On the one hand, cooperation in cheating may reduce enforcement by reducing the likelihood of whistleblowing (Rullo et al., 2024). On the other hand, in the case of enforcement, the consequences may appear less severe, due to the diffusion of responsibility within the group (Zanetti & Butera, 2023).

3.1.3 Intergroup processes in dishonesty

Dishonesty occurs not only between different individuals in a group, but also across group boundaries. Although in real life, dishonest acts often involve multiple groups (e.g., bribery across national borders), research on this topic is still scarce. In fact, although an emerging stream of research on collective dishonesty exists (for a meta-analysis, see Leib et al., 2021), intergroup dishonesty has only recently been studied. In a large-scale multinational study by Dorrough et al. (2023), participants from different countries interacted in a bribery game in mixed-national pairs. Results show people offer disproportionately more bribes to interaction partners from countries with a high (vs. low) reputation for foreign bribery, as measured by macro-level indicators of corruption perceptions (e.g., the BPI). People largely share country-specific expectations about the level of acceptance of bribery in other countries. However, these expectations are negatively correlated with actual levels of bribe-taking, suggesting shared but inaccurate stereotypes about bribery propensities. Findings also suggest more than one's own national background and the national background of one's interaction partner shape people's willingness to engage in bribery, highlighting the importance of expectations and stereotypes for dishonest behavior. In this research, the decision context was highly relational with low enforcement (see Fig. 2). Future research can implement different punishment institutions (i.e., high enforcement) and test their effectiveness in combating dishonest behavior in the form of bribery and beyond.

In addition, the role of gender in corruption research has received considerable attention (Kubbe & Merkle, 2022). However, similar to the national group membership of the opposing party, the gender composition of the group with which one interacts has not yet been examined. This task is left for future research. The same is true for other instances of dishonest behavior (e.g., lying, cheating) and other category memberships (e.g., high-status vs. low-status groups).

Psychological science tends to test the generality of phenomena, but showing how phenomena vary as a function of context is equally important. An open research question in (dis)honesty is the role of context and how a culture of cheating manifests itself in different ways at work, in sports, at school, in the arts, and so on. Indeed, research has shown a culture of cheating (vs. not cheating) affects moral disengagement (Zanetti & Butera, 2024).

Showing how routinized cheating leads to reduced feelings of guilt, and how highly competitive environments—for example, in industry or science—can lead to collective moral disengagement (Bandura, 2002)—for example, in the form of beliefs that "the end justifies the means"—would be interesting.

3.2 Relationship features

3.2.1 Social norms

Social norms form the backbone of our societies, guiding both selfish and cooperative behavior and enabling the formation of social bonds. These norms are often maintained through credible enforcement mechanisms and are crucial in influencing individual decisions, particularly in the context of deviant behaviors such as lying and dishonesty. How do social norms affect and are affected by lying?

Norms are properties of an individual's relevant reference group, which are people who are important to the decision-maker in a particular situation (Bicchieri, 2006; Krupka & Weber, 2013; Fehr & Schurtenberger, 2018; Bicchieri & Dimant, 2019; Cohn et al., 2019, 2022). People often observe and imitate these reference groups, interpreting their intentions and sometimes justifying their own actions and the underlying social norms in self-serving ways (Bicchieri et al., 2023). This process highlights the importance of understanding how norms and compliance evolve, particularly in settings where individuals can choose which norms to learn and follow (Dimant et al., 2024).

Peer behavior influences one's choices through at least one of two channels that follow from Bicchieri (2006): empirical expectations (what others do) and normative expectations (what others approve of). For example, when people know their peers also lie or engage in deviant behavior, they are more likely to behave in a similar way, expressing a form of conditionality (Bicchieri et al., 2022). Sometimes, however, people are unable to fully observe their peers or are not observed by their peers. Such environments are characterized by the anonymity of actions, which hinders the enforcement of social norms and thus facilitates the spread of dishonesty (Dimant et al., 2020; Kajackaite & Gneezy, 2017).

Importantly, the lack of observability and norm enforcement creates moral space through which deviance can spread (Bolton et al., 2021). As noted above, people use self-interested justifications to determine how much to lie and strive to maintain a positive self-concept even when engaging in dishonest behavior (Bicchieri et al., 2023). When individuals had multiple opportunities to roll a die and report the outcome, they were more likely to lie by reporting a favorable but unobserved outcome. These findings suggest lying feels more justified when individuals can convince themselves they are reporting a possible outcome rather than fabricating it entirely (Shalvi et al., 2011). At the same time, some wiggle room remains in terms of what social information one is exposed to. In experiments in which participants were allowed to choose their sources of information about prevailing social norms, they showed a bias toward more lenient information (Dimant et al., 2024). This bias was stronger among those who had already lied, suggesting individuals actively seek out information that justifies their dishonest behavior. This selection bias in information acquisition can have significant implications for the formation and influence of social norms (Dimant et al., 2024; see also Maggian & Villeval, 2016).

Finally, social norms can be a fragile construct. Social-norm enforcement is essential for maintaining social order and reducing the occurrence of dishonesty (Buckenmaier et al., 2021). Norm-enforcement behavior can be influenced by norm nudges, which can change individuals' perceptions of what behaviors are acceptable. For example, providing normative information about the inappropriateness of lying can increase the perceived inappropriateness of the behavior, thereby increasing the likelihood of norm enforcement through punishment (Bicchieri et al., 2021; Dimant & Gesche, 2023). Similarly, Capraro and Vanzo (2019) demonstrated the use of morally loaded language can influence people's honesty in decision-making contexts, highlighting the power of framing in ethical behavior.

3.2.1.1 Reducing dishonesty via norm nudging

Norm nudging leverages social expectations to change behavior by influencing what people believe others will do or approve of (Bicchieri & Dimant, 2019). This method has been shown to be effective in a variety of related contexts, such as tax compliance and the enforcement of honesty norms (Hallsworth et al., 2017). For example, informing taxpayers about the high compliance rates of their peers increased tax compliance rates. These examples illustrate norm nudging can be a powerful tool for promoting honesty.

An important aspect that determines the effectiveness of norm nudging is the relevance of the peer group and the context in which the behavior occurs. For example, one study found hotel guests were more likely to reuse towels when they were informed that previous guests in their room had done so. This finding suggests that the closer the reference group is to the individual's context, the more effective the norm nudge will be

(Goldstein et al., 2008). However, norm nudges do not always produce positive behavior change, and the effectiveness of such interventions can vary significantly depending on how the norm information is framed (Bicchieri & Dimant, 2023). In a study examining the effects of norm nudges on lying behavior, Dimant et al. (2020) found framing the norm message positively (e.g., "the majority did not cheat") or negatively (e.g., "a minority did cheat") had no significant effect on participants' dishonesty and was indistinguishable from a no-information control condition. This finding suggests norm nudges may not always be sufficient to change existing norms, especially when those norms are already well established (Dimant, 2024; Gelfand et al., 2022).

In sum, social norms play a critical role in guiding behavior and maintaining social order. Understanding how norms influence deviant behavior and developing effective interventions can foster a more honest and cooperative society. Although existing research has provided valuable insights into the role of social norms in deviant behavior, several gaps remain. For one, future research should examine how norms form endogenously through the selection of information in social contexts and how this process influences behavior. In addition, more research is needed to examine the long-term effects of norm nudges and the conditions under which they are most effective. This examining includes how different framing strategies and the presence of group identity influence the effectiveness of these interventions.

Finally, in addition to direct effects on social norms and behavior, the relationship between deviance, lying, and political polarization is of growing importance in today's political climate. Research suggests political polarization may exacerbate the spread of deviant behaviors such as lying, because individuals may selectively seek out and trust information that is consistent with their political beliefs, thereby reinforcing dishonest behaviors within their in-groups (Dimant et al., 2024; Dimant, 2024). This phenomenon is exacerbated by the existence of echo chambers and filter bubbles, where exposure to homogeneous sources of information can reinforce partisan views and justify dishonest actions toward out-groups (Bursztyn et al., 2020). Understanding these dynamics is critical to designing interventions that promote honesty and cooperation across political divides. Importantly, to successfully harness the power of social norms and norm nudges to curb dishonesty, we need to develop new measurement approaches that capture the nuances of social norms. Existing measurement approaches (e.g., Bicchieri & Chavez, 2010; Krupka & Weber, 2013) are useful in capturing the average or most common behavior/opinion in a given context. However, the new generation of social-norm-elicitation techniques (e.g., Dimant, 2023; Panizza et al., 2024) also capture more nuanced aspects of social norms, such as tightness, looseness, and polarization.

Taken together, understanding the shape of social norms, the mechanisms underlying norm adherence, and their enforcement in different contexts can help design more effective interventions to promote sustained honesty (Dimant & Shalvi, 2022).

3.2.2 Trust

The previous sections show that although humans often portray themselves as moral beings, their actions often contradict this self-image. As we have seen, people engage in dishonesty, form fraudulent alliances, and disregard social norms. However, the extent of such behavior is influenced by numerous factors, including individual differences, the likelihood of being caught, and cognitive factors such as mental capacity and the (lack of) energy required to deceive. Given the prevalence of dishonest behavior, how can people still trust each other?

Trust is defined in terms of capabilities (the belief that the trusted individual is competent to perform a task) and encapsulated interests (the expectation that the trusted individual will consider the interests of others). Seligman (2000) described trust as "a kind of belief in the other's benevolence, given the opacity of the other's intentions and calculations." Similarly, Russell Hardin (2002) defined it as the expectation of "encapsulated interest," that is, the understanding that acting on one's behalf is in the trusted person's interest. This logic sounds rather "sophisticated. In fact, some trust decisions that people have to make in their lives are made after extensive thought and careful consideration. Most of the time, however, we make trust decisions in a split second, based on gut feelings rather than rationality (Li et al., 2022; Snowden et al., 2022).

Using a trust game, Engelmann and Herrmann (2016) showed chimpanzees place more trust in their partners when interacting with friends than with non-friends. Based on this and a handful of other studies, we can conclude that trust within close relationships is not a uniquely human trait. But what about trust in complete strangers? Observational studies in the wild show chimpanzees and our other close relatives, namely, bonobos, behave quite differently in encounters with unfamiliar conspecifics, with the former generally being more aggressive than the latter (Hare et al., 2012), although other socio-ecological factors (e.g., gender, ingroupoutgroup, available resources) somewhat blur this species difference (Mouginot et al., 2024). Studies of human hunter-gatherers show that, compared with chimpanzees living in the same Congolese rainforest, humans travel greater distances and trade with other groups (Jang et al., 2019). However, their interactions with strangers are minimal compared with people in WEIRD countries (Westernized, Educated, Industrialized, Rich, and Democratic).

Humans naturally organize themselves into groups and subgroups, but our WEIRD society has become extremely layered due to globalization and digitization. Permissions, registrations, benefits, rules, acts, laws, inspections, and repercussions, but also the sharing of experiences via social media, ensure people and institutions strive to maintain their reputations. In the Netherlands, for example, a daycare provider must have completed training as a childcare worker and clear a criminal record check before being allowed to care for other people's children. The complexity of our society is strange, but on an evolutionary timescale, it has helped reduce aggression and foster trust and large-scale cooperation, even with strangers (Pinker, 2012). Research on non-human primates and non-WEIRD societies is extremely scarce. But if we are to better understand the precise societal, cultural, ecological, and evolutionary drivers of trust, comparative research is essential. Such a line of research would open up opportunities for cross-disciplinary collaborations between psychologists, anthropologists, ecologists, primatologists, and behavioral economists.

The above has provided some explanations for why humans, although we often cheat each other, still trust on a large scale. The question of why we feel more comfortable sitting next to person A on the train than person B remains open. A crucial factor here is what we see in their faces. To explore which facial features inspire trust, the researchers used a technique called "reverse correlation" (Dotsch & Todorov, 2012). Participants were shown the same grayscale face on a computer 1000 times, each time with a different noise filter, namely, a random pattern of pixels. They saw two faces at the same time and had to decide which looked more trustworthy, the left or the right. The pixel patterns changed the faces slightly each time. Overlaying all the pixel patterns deemed trustworthy created a baby face with a round jawline, large pupils under light eyebrows, a high forehead, and slightly upturned corners of the mouth (Dotsch & Todorov, 2012; Frisanco et al., 2023). Conversely, superimposing the untrustworthy pixel patterns resulted in a grumpy blockhead, the stereotypical "untrustworthy person.

Follow-up research showed people were more likely to trust people with a baby face. In fact, people need only 33 ms to categorize a face as "trustworthy" or "untrustworthy" (Todorov et al., 2009). But the story does not end there. Although we might be more inclined to entrust our child to a person with a childlike face and friendly expression, we tend to trust politicians with features that convey dominance and competence: a square jaw, eyes that are not too wide, high cheekbones, and preferably some attractiveness (Laustsen & Petersen, 2018). Research shows these facial features significantly influence election outcomes. Our primate brains often struggle in our modern society. The associations, biases, and stereotypes we consciously or unconsciously hold make us less likely to trust a female politician or a male daycare worker. These undesirable evolutionary "leftovers" are reinforced by what we see and become accustomed to on a daily basis. Importantly, we need to remember that although the way humans live today is unique, our genome evolved in small, close-knit groups where encounters with strangers were rare.

In addition to static facial features, emotional expressions are also critical in assessing trust. Many studies have shown people with a smile are more trusted than those with a frown, and that even subtle cues such as blushing (Dijk et al., 2011), pupil dilation (Kret & de Dreu, 2019), or direct gaze (Kreysa et al., 2016) can make one appear more trustworthy. Emotional expressions are often imitated by others, and people even synchronize their physiological states. Synchronization helps predict each other's behavior, which is why various studies have shown mimicry and synchrony promote trust (Kret & Akyüz, 2022). After decades of silence following Levenson and Gottman's (1983) seminal paper on physiological coupling in couples, new statistical methods and affordable equipment have revitalized the field of social neuroscience. Research has shown people mimic pupil size and that mimicking pupil dilation has prosocial consequences, including an increase in trust (Kret et al., 2015). By combining eye tracking, psychophysiological measures, and fMRI, researchers revealed the importance of the theory of mind network for this link between pupil mimicry and trust. Every year, new products, such as wearables that measure physiological responses or brain activity, enter the market. These innovations allow researchers to study the precursors of trust decisions during real-life interactions, with a focus on physiological synchrony and hyper-scanning. One of the most challenging and exciting research questions we face is understanding how the intricate choreography of emotional expressions between interaction partners influences mutual trust.

3.2.3 Gossip

Our framework (Fig. 2) highlights the need to explore understudied areas related to enforcement and the need to examine decision contexts that are relational rather than solitary. Research on gossip relates to both of these recommendations. Gossip is defined as communication between a sender and receiver about an absent person (i.e., the gossip target; Dores Cruz et al., 2021). When the opportunity to engage in dishonest behavior arises, the possibility that others will gossip about such behavior implies, in the terminology of our framework, that the behavior can be detected and punished by others (i.e., gossip can be viewed as a form of coercion) and that the decision context involves others (i.e., the presence of gossip makes the context relational). Taken together, reviewing research on the influence of gossip on (dis)honesty is valuable here.

Gossip allows individuals to learn about the behavior of others without directly observing them, and as such, researchers have argued it is an essential mechanism explaining human cooperation (Nieper et al., 2022; Wu et al., 2016). Through gossip, receivers can learn whom to trust and cooperate with and whom to avoid or punish (Dores Cruz et al., 2021; Sommerfeld et al., 2007), thus allowing for indirect reciprocity (cooperating with those who cooperated with others; Nowak, 2006) and partner selection (choosing group members who are known to be reliable partners, Capraro et al., 2016). Therefore, for potential gossip targets, the possibility of being gossiped about should lead to reputational concerns (i.e., concerns about others' evaluations of one's qualities or characteristics; Caldwell, 1986), which in turn should make them more likely to behave in ways that lead to a positive reputation (e.g., Beersma & van Kleef, 2011; Wu et al., 2016).

Linking gossip to honesty, the possibility of being the target of gossip should, in principle, lead people to behave more honestly due to reputational concerns. To examine the functioning of gossip as an enforcement mechanism, Nieper et al. (2025) used the aforementioned die-rolling task. Participants who received reports about the behavior of dice rollers and were able to gossip about them to other participants largely portrayed gossip targets as honest or dishonest, consistent with the dice roll reports they received: die roll averages were strongly correlated with gossip describing the gossip target as dishonest. Moreover, gossip recipients used gossip as a valid source of information: those portrayed as dishonest were trusted less than those portrayed as honest. In addition, and central to the framework presented here, dice rollers' honesty increased when they knew the participant observing their dice rolls could gossip about them to another participant with whom they would interact later in the experiment, whereas mere observation without the possibility of gossiping did not increase honesty. The effect of (potential) gossip on honesty was partially mediated by reputational concern.

Although these findings suggest the (possibility of being the target of) gossip can have a positive effect on honesty, and thus on the success of gossip as an enforcement mechanism, results from the second study by Nieper et al. (2025) show this effect is limited to situations in which gossip receivers can influence the outcomes of gossip targets. Specifically, the ability of observers of dice rolls to gossip to someone who could not influence the participants' monetary outcomes in a follow-up task did not increase the honesty of the dice rollers, whereas it did when observers were able to gossip to someone who could directly influence the dice rollers' monetary outcomes in that task. Similarly, Huber et al. (2023) found no effect of a manipulation that combined observability and non-consequential gossip on honesty in the dice-rolling task. Thus, the possibility of being the target of gossip appears to affect honesty only when gossip is directed to receivers who can impose consequences on potential gossip targets, suggesting gossip is particularly effective as an enforcement mechanism when it can lead to actual punishment of the gossip target by gossip receivers.

Interestingly, in addition to acting as an enforcement mechanism against dishonesty, gossip itself can reflect dishonesty. People can lie in their gossip and this is more likely to happen when it is personally beneficial (Dores Cruz et al., 2024; Peters & Fonseca, 2020). Therefore, an important question for future research is to what extent gossip needs to be accurate in order to stimulate honest decision-making. If decision-makers know gossip about them may be inaccurate, and if they know gossip recipients know this, will they still behave more honestly? An agent-based simulation study by Testori et al. (2022) suggests such inaccuracy can counteract the positive effects of gossip on group cooperation via indirect reciprocity and/or partner selection. Although research has shown that even when receivers knew gossip statements may contain inaccurate information, the gossip still affected their evaluations of targets (Baum et al., 2020), whether and how (potentially) inaccurate gossip affects the behavior of targets is unclear.

Related to the potential effects of gossip inaccuracy, the influence of the interdependence structure between the three members of the "gossip triad" (the sender, the target, and the receiver) represents another interesting avenue of research. Although the results discussed above show the possibility of being the target of gossip increases honesty, we do not currently know whether it does so under different interdependence structures. Results from the second study by Nieper et al. (2025) show gossip increases honesty only when the outcomes of gossip targets depend on the decisions of gossip receivers, but the effects of other patterns of interdependence in the gossip triad, such as when an observer who can gossip about one's behavior would benefit from lying, or when a gossiping observer is an ally or an enemy of the gossip receiver, are still unclear.

In a situation where a gossiper could benefit from a target's dishonesty (see Weisel & Shalvi, 2015), negative gossip might portray the target as dishonest when they lie, but as uncooperative when they behave honestly. Thus, if targets are primarily concerned about their reputation as honest people, gossip in this situation should lead to more honest decisions, but if targets are primarily concerned about their reputation as good cooperators, gossip should lead to more lies. Nieper et al. (unpublished manuscript) investigated how gossip affects the behavior of potential targets in this situation. Their preliminary results, based on Bayesian statistics, showed a null effect of gossip. However, they found the effect of gossip was moderated by the target's personality, such that gossip led to more honest reports for targets who scored low on honesty-humility. An interesting possibility that future studies could explore is that individuals with nefarious goals may use gossip about others' honesty versus dishonesty as a way to identify potential partners for collaborative cheating. Thus, although gossip generally appears to promote honesty, it may ironically enable dishonesty to the extent that it allows individuals to identify future partners in crime.

Finally, many studies examining the effects of gossip on target behavior have examined the possibility or threat of gossip rather than becoming the victim of gossip (see, e.g., Beersma & van Kleef, 2011; Fehr & Sutter, 2019). Research has shown the experience of being the target of gossip leads to reduced feelings of social inclusion and withdrawal (Martinescu et al., 2021). Testing the effects of actual, rather than anticipated, gossip on honesty would therefore be interesting.

3.3 Individual differences

3.3.1 Personality, beliefs, and justifications

Research has firmly established notable interindividual differences in whether and to what extent people engage in unethical behaviors such as dishonesty, and that this variation can be explained by personality traits (for a review, see Hilbig, 2022). In particular, the honesty-humility dimension from the

HEXACO model of personality structure (Heck et al., 2018; Hilbig & Zettler, 2015; Zettler et al., 2020) and the common core of so-called "dark" traits (e.g., narcissism or psychopathy), the "dark factor of personality" (hereafter "D"; Moshagen et al., 2018; Moshagen et al., 2020). These two traits not only predict lying in anonymous, individual settings, but have also been shown to predict how consistently individuals tell the truth or lie in such contexts over time (Thielmann et al., 2024). Similarly, as noted in the introduction, research has well established that people who lie rely on self-preserving justifications to avoid threats to their moral self-image.

Until recently, however, these two approaches to understanding dishonesty—personality traits and justifications—have typically been studied in isolation. With the goal of integrating these lines of research, Hilbig et al. (2022) proposed that certain beliefs associated with the D factor of personality explicitly defined as "the general tendency to maximize one's individual utility—disregarding, accepting, or maliciously provoking disutility for others—are accompanied by beliefs that serve as justifications" (Moshagen et al., 2018, p. 657, emphasis added) promote the subjective justification of disutility, essentially acting as an accelerant that operates on the underlying motivation, namely, the trait. In other words, individuals who are more motivated and more likely to lie (those high in D) are expected to hold beliefs that allow them to construct justifications for dishonesty and thereby mitigate threats to their moral self-image.

Prominent among such beliefs inherent in D are those related to superiority, special status, and outstanding deservingness, which is relatively unsurprising given that traits such as grandiosity, narcissism, psychological entitlement, or self-centeredness can all be considered mere manifestations of D (Hilbig et al., 2023; Scholz et al., 2024). However, D is also associated with a much broader set of beliefs that may justify dishonesty. For example, those high in D hold hierarchy-related beliefs (especially social-dominance orientation and/or authoritarianism; Moshagen et al., 2024) and relativism-related beliefs (such as normlessness or negative reciprocity norms), both of which operate on injunctive social norms (called normative expectations in Section 3.2.1), the former justifying dishonesty in terms of its necessity (for the in-group or as demanded by some authority) and the latter in terms of ambiguity and conditionality, and thus the absence of a norm. Most importantly, however, D is strongly associated with various distrust-related beliefs (such as a competitive worldview or cynicism, i.e., "everyone does it") that invoke a descriptive social norm (called empirical expectations in Section 3.2.1) and thus justify dishonesty by its prevalence (e.g., Gächter & Schulz, 2016; López-Pérez, 2012; Schultz et al., 2007; see Section 3.2.1). Experiments (Hilbig et al., 2022) have shown individuals high in D are more likely to predict others will be dishonest (i.e., cynicism, measured in an incentive-compatible manner; i.e., participants received bonus payoffs for correct predictions about others' dishonesty). In turn, these individuals are more likely to judge dishonesty as justifiable, and ultimately more likely to engage in dishonesty themselves, supporting the notion that the (cognitive) mechanism by which personality is expressed (see Section 3.3.2) involves the construction of justifications from beliefs.

In related work using a complex moral-dilemma paradigm, Ng et al. (2022) found individual differences in deontological norm adherence negatively predicted cheating in a coin-toss paradigm, regardless of whether cheating was for personal gain or the greater good. By contrast, individual differences in utilitarian welfare maximization did not predict cheating regardless of whether it served personal gain or the greater good. Although not of primary interest, the two studies also included the honesty-humility subscale of the HEXACO inventory, which negatively predicted cheating for personal gain but not cheating for the greater good.

Perhaps one of the most fundamental open questions in this regard is essentially ontological: do individuals who are initially inclined toward dishonesty successively adopt increasingly strong and varied specific beliefs (e.g., cynicism or social-dominance orientation) to protect their moral self-image, or do individuals initially hold or acquire (e.g., through experience and exposure) certain beliefs that then gradually incline them toward unethical behavior? Whether and which social contexts and experiences shape these beliefs, especially those related to the honesty of others and the likelihood and effectiveness of sanctions, remains unknown (see Fig. 2). For example, cynical beliefs include the agreement that "most people aren't really honest for a desirable reason; they're afraid of getting caught" (Chowdhury & Fernando, 2014), and relativistic beliefs assert that "if something works, it doesn't matter whether it's right or wrong" (Kohn & Schooler, 1983). Thus, the very beliefs that link personality traits to dishonest behavior appear to be directly related to the two major dimensions identified in the framework above. Thus, an important task for future research is to determine not only how the likelihood of sanctions and the social context shape dishonest behavior (as in Fig. 2), but also how related experiences play into the very beliefs that link personality traits to dishonest behavior.

3.3.2 Honesty and personality dynamics

In personality science, an exciting recent line of research on honesty has emphasized personality dynamics (Furr & Funder, 2021; Rauthmann, 2021). Whole Trait Theory, for example, posits that traits include both consistent patterns of behavior (a descriptive aspect) and the underlying psychological mechanisms (an explanatory aspect) that cause these patterns (Fleeson & Jayawickreme, 2021). A dynamic perspective recognizes people regulate honesty based on situational contexts and individual goals, emotions, and cognitive processes.

By considering honesty from this perspective, dynamic personality accounts acknowledge that although some individuals may be inclined to be honest, due to stable personality traits, their honesty may vary meaningfully depending on specific contexts. For example, a person may be generally honest but lie when he or she perceives a threat to his or her self-image or relationships. Such variability arises from psychological processes such as motivations, values, and beliefs that are activated in different contexts. Thus, dynamic accounts of personality provide a perspective for understanding how consistent patterns of honesty emerge over time while also accounting for flexibility in honesty across contexts in daily life.

The framework outlined in Fig. 2 can be viewed through the lens of personality dynamics. First, enforcement can be broadly interpreted as "the perceived likelihood of one or more consequences occurring and triggering a sanction" and thus in terms of goal-directed decisions. Actors' beliefs about the likelihood of lie detection and subsequent enforcement appear to be most relevant when actors worry about undesirable interpersonal or relational consequences that may occur if others detect their dishonesty. This view can be extended in at least two ways. First, lie detection might produce desirable consequences (e.g., being "caught" lying to help another person might produce appreciation and praise). Second, dishonesty may produce undesirable consequences that are not interpersonal or relational. For example, if an agent views dishonesty as immoral, her dishonest behavior might produce the undesirable intrapersonal consequence of seeing herself as immoral, even if no one else detects it.

This extended view suggests that when deciding whether to lie in real-life contexts, actors are likely to consider the likelihood that doing so will produce both (a) ≥ 1 desired consequence and (b) ≥ 1 undesired consequence. When viewed as goal directed (consistent with dynamic approaches such as Whole Trait Theory, which views flexibility in moral behavior as intentional), decisions to act (in)honestly arise from weighing the likelihood of competing outcomes and the relative desirability of those outcomes. Greater likelihood of detection thus means one or more consequences (most likely, but not only, undesirable interpersonal consequences) are more likely to occur.

Second, the dynamic approach can also frame the "decision setting" facet of Fig. 2 broadly in terms of attention to the relational context of the honesty-relevant situation. This approach emphasizes the importance of different contextual factors in shaping (in)honest behavior, consistent with dynamic accounts of personality. Varying relational factors can manifest themselves in a variety of ways, including variation in the types of consequences that are possible, changes in the relative importance of consequences, and shifts in the likelihood that particular consequences will occur.

These two facets call for the study of honesty through the lens of goal-oriented personality dynamics outside the laboratory. Goals, which reflect important differences among people but are also highly contextualized, are an important part of dynamic approaches to personality (Jayawickreme et al., 2017). Particular goals (e.g., financial reward, avoidance, of interpersonal sanction) are activated more frequently in some situations than in others. Similarly, certain goals are more highly valued, viewed as more achievable, and perceived as more possible for some individuals than for others. To fully understand an individual's (dis)honesty, we need to understand both the situational factors that shape the availability and achievability of (dis)honesty-relevant goals and the personality factors that shape which goals are considered desirable and attainable (by some people).

This perspective has both methodological and conceptual implications. Methodologically, insights into (dis)honesty will emerge from studying individuals' honesty-relevant goals, perceptions, beliefs, and behaviors as they move from situation to situation. An extremely useful approach to do so is through intensive assessments such as experience sampling methods (ESM, e.g., Meindl et al., 2015), which provide the opportunity to efficiently and quickly "observe" many individuals in many real-life situations. Indeed, consistent with the dynamic approach, recent work has used intensive measurement to assess the frequency of honest behavior and the psychological processes, such as motivations and beliefs, that drive that behavior. One challenge here (depending on how dishonesty is defined) is a potentially low base rate of dishonesty in daily life, or at least dishonesty that is more than trivial (e.g., something more than "I'm feeling fine today, thank you" when, in truth, one is feeling slightly less fine than usual that day). If meaningful or significant dishonesty is relatively rare in everyday life, researchers will need to consider which types of dishonesty are of interest and (potentially) how to study a phenomenon with a low base rate.

Conceptually, the dynamic perspective highlights several important questions. First, and most generally, what are the mechanisms underlying moral traits such as honesty (Sun & Smillie, 2024)? Second, what kinds of real-life goals are generally facilitated by honesty, by dishonesty, or potentially by both (depending on the situation)? Third, to what extent do people differ in their valuing, believing in their ability to achieve or avoid, or perceiving the availability of certain (dis)honesty-relevant goals? Fourth, to what extent can person-level differences in (dis)honest behavior be explained by goal-related dynamics?

3.3.3 Care for honesty vs. pro-sociality

As noted above, one of the most established trait predictors of honesty is honesty-humility from the HEXACO personality model, which represents "the tendency to be fair and genuine in dealing with others" (Ashton & Lee, 2007, p. 156) and thus the link between honesty and prosociality. An open research question in this context, which has recently received increasing attention, is whether the relationship between honesty-humility and honest behavior remains robust when the latter conflicts with prosociality, that is, when being dishonest is associated with increasing the welfare of others, that is, being prosocial. As defined above, honesty-humility involves both fairness and sincerity; thus, individuals high in honesty-humility should be torn between honesty and prosociality whenever the two behaviors remain in conflict. In a recent attempt to address this question, Thielmann et al. (2024) manipulated who benefited from an individual's dishonesty: the individual, a "non-needy" other (e.g., another participant with an average household income), or a "needy" other (e.g., another participant with a household income below the poverty line). In the latter two conditions, dishonesty is purely prosocial because only another person benefits from an individual's lie. In turn, helping someone in need arguably involves a stronger moral imperative than helping a non-needy stranger; thus, individuals motivated by prosociality should feel particularly drawn to dishonesty when they can help a needy other. However, regardless of how morally imperative lying was, the positive relationship between honesty-humility and honesty remained robust, suggesting dispositionally honest individuals prioritize honesty over prosociality when the two conflict.

Evidence on the relationship between honesty-humility and collaborative dishonesty paints a similar picture. In one study, honesty-humility showed a negative relationship with collaborative dishonesty, regardless of whether individuals faced a fully honest or dishonest counterpart (Ścigała et al., 2019). Similarly, honesty-humility predicted honest behavior in the Unethical Loyalty Game (Seidl et al., 2024; Thielmann et al., 2021). In this game, individuals in the role of witness can lie to cover up a transgressor's dishonesty and thus be unethically loyal. Unethical loyalty confers a benefit on the transgressor (as well as the witness) and can therefore be considered prosocial. Abstaining from unethical loyalty (i.e., being honest about the transgressor's dishonesty), on the other hand, leaves the transgressor (and the witness) empty-handed. In both a laboratory (Thielmann et al., 2021) and an online (Seidl et al., 2024) version of this game, honesty-humility was negatively related to unethical loyalty, meaning individuals with higher levels of honesty-humility were more likely to refrain from lying to improve another person's outcome.

Overall, the evidence suggests dispositionally honest individuals—as reflected in trait levels of honesty-humility—value honesty over prosociality. On the other hand, in most studies examining individual differences in prosocial lying, the prevalence of dishonesty was above zero even among individuals high in honesty-humility (see also Ścigała et al., 2020). In other words, even people who tend to be honest sometimes lie when doing sot benefits others, and this tendency seems to be especially true when lying has a collaborative aspect (see, e.g., Thielmann, Hilbig, Klein, et al., 2024). Future research is needed to provide a more systematic understanding of how different individuals weigh honesty versus pro-sociality in different contexts (e.g., solitary vs. relational settings; see Fig. 2) as a function of their personality traits. Moreover, all previous studies implemented paradigms in which lying remained anonymous to others, except for the interaction partner (in some cases). Thus, in line with the above framework (Fig. 2), another avenue for future research might be to examine how personality manifests in prosocial lying when enforcement is high, for example, because dishonesty can be detected (and potentially punished) by uninvolved parties.

3.3.4 Gender differences

Since the seminal observation that men are more likely than women to tell self-serving lies in everyday contexts (DePaulo et al., 1996), behavioral scientists have extensively used economic games to explore potential gender differences in the propensity to lie for personal gain (Kennedy & Kray, 2022).

Early research by Dreber and Johannesson (2008) confirmed that, on average, men lie more than women. This finding has been replicated in several studies (Babin & Chauhan, 2024; Capraro et al., 2019; Conrads et al., 2013; Elaad et al., 2024; Friesen & Gangadharan, 2012; Muñoz García et al., 2021; Ruffle & Tobol, 2017). However, other studies have not found significant gender differences in lying (Childs, 2012; Ezquerra et al., 2018; Lundquist et al., 2009; Alfonso-Costillo et al., 2022). One study even found women lie more than men (Ruffle & Tobol, 2014).

Given this variability, considering meta-analyses that aggregate findings across multiple studies to provide a more comprehensive picture is useful. Capraro (2018) conducted a meta-analysis of 4173 decisions in a sender-receiver game and found men were more likely than women to tell selfish lies. A subsequent, larger meta-analysis by Gerlach et al. (2019), which included over 40,000 observations from the sender-receiver game, matrix task, die-roll task, and coin-flip task, confirmed this trend. They found men are generally more dishonest than women, with an estimated 4-percentage-point difference in honesty. The experiments included in these meta-analyses were primarily conducted in Western societies. However, a recent working paper suggests gender differences in honesty may be culturally shaped. Whereas women are more honest than men in Western societies, this pattern does not hold in several non-Western contexts, partly due to differing social norms (Graf et al., 2024).

3.3.4.1 Gender difference in self-serving dishonesty: Other-regarding preferences or intrinsic cost of lying?

Why are women more honest than men? Is the reason that they have a higher intrinsic cost of lying, reflecting a greater moral aversion to dishonesty? Or is the reason that they are less selfish, that is, they care more about the other person's or the experimenter's payoff?

Dictator game studies have shown women tend to give more than men (Brañas-Garza et al., 2018; Doñate-Buendía et al., 2022; Engel, 2011; Rand et al., 2016). This finding suggests women may indeed be more concerned with the payoffs of others, which may partly explain their lower levels of self-serving dishonesty. Consistent with this view, one study found social-value orientation mediates gender differences in lying (Grosch & Rau, 2017).

To examine whether the intrinsic cost of lying also plays a role, consulting the literature on gender differences in lying that benefits another person, known as white lies, is useful (Erat & Gneezy, 2012). White lies can be categorized into altruistic white lies, which help another person at a cost to the liar, and Pareto white lies, which benefit both the liar and the receiver. If women have a higher intrinsic cost of lying, we should observe higher levels of honesty among women in these contexts as well.

In the case of altruistic white lies, initial evidence suggests women tend to tell more altruistic white lies than men (Erat & Gneezy, 2012). However, Biziou-van-Pol et al. (2015) did not replicate this finding and instead found women tell fewer altruistic lies than men. Another study found no gender differences in lying when the benefit goes to a charity (Babin & Chauhan, 2024). A meta-analysis of 2940 decisions in a sender-receiver game found women are significantly less likely than men to tell altruistic white lies (Capraro, 2018). These findings suggest women's lower levels of self-serving dishonesty cannot be attributed solely to other-regarding preferences but is partly due to a higher intrinsic cost of lying.

In the context of Pareto white lies, Erat and Gneezy (2012) found that, on average, men lie more than women. However, Cappelen et al. (2013) did not replicate this finding and did not observe gender differences. An internal meta-analysis of 1615 decisions found a marginally significant tendency for men to tell more Pareto white lies than women (Capraro, 2018). However, this meta-analysis is limited by its small sample size and internal nature, which reduces its statistical power.

Research on white lies suggests gender differences in self-serving lying cannot be explained solely by gender differences in other-regarding preferences. Two pieces of evidence further support this view. D'Attoma et al. found women in five countries tended to report more income than men in tax-compliance experiments, even though no consistent gender differences in prosociality in these countries were present. Leib et al. (2021) conducted a meta-analysis of cooperative dishonesty and found groups with higher proportions of women tended to be more honest.

In sum, existing research suggests gender differences in self-serving lying are likely due to a combination of factors: men tend to be more self-interested and have a lower intrinsic cost of lying than women.

3.3.4.2 Heterogeneity in gender differences: Probability of detection and group interactions

The observed gender differences in self-serving lying hold on average, but considerable heterogeneity exists. Given the conceptual framework in Fig. 2, an important potential source of heterogeneity may be enforcement mechanisms, specifically the likelihood that a lie will be detected.

Theoretically, the risk of detection might reduce lying especially among risk-averse individuals. Because women are generally more risk-averse than men (Byrnes et al., 1999), the risk of detection might increase honesty more among women. However, experimental evidence is scarce. To our knowledge, no studies have specifically examined gender differences by experimentally manipulating the probability of detection. One study examined gender differences at a fixed detection probability of 0.5 and found women were 9% point more honest than men (Lohse & Qari, 2021). Given the overall 4-percentage-point gender difference in lying estimated by Gerlach et al. (2019), this finding could, at best, be seen as preliminary evidence supporting the theoretical prediction that detection probability may increase gender differences in lying. Future research should explore this possibility more thoroughly.

Other important sources of heterogeneity lie in the relationships between actors. As discussed in the previous section, whether gender differences exist in the telling of Pareto white lies remains unclear. Investigating this possibility further is an important direction for future work. Another important aspect to consider is group interactions. Previous research has shown groups are more likely to lie than individuals, primarily due to two mechanisms: communication and learning about norm compliance. Communication can expose individuals to arguments in favor of norm violation, which can change expectations about norm compliance and subsequently affect behavior (Kocher et al., 2018). These mechanisms may affect women and men differently. Because the intrinsic costs of lying are generally higher for women, group interactions that reduce these costs could lead to increased dishonesty, particularly among women. In support of this view, Kocher et al. (2018) found group interactions increased dishonesty primarily among women. Furthermore, Muehlheusser et al. (2015) found all-female groups lie less frequently than all-male and mixed-gender groups, with the latter two groups exhibiting similar levels of dishonesty. Taken together, these findings suggest a possible explanation: group interactions may increase the likelihood that women will lie by exposing them to men's arguments in favor of lying. Future research could test this hypothesis.

3.4 Cognitive processes

3.4.1 Ethical dissonance

Ethical dissonance is the psychological tension between the temptation to act unethically for personal gain and the desire to maintain a moral self-image (Barkan et al., 2015). This unique form of cognitive dissonance (Elliot & Devine, 1994; Festinger, 1962) involves violations of absolute moral values that pose a significant threat to self-integrity. Unlike typical dissonance, ethical dissonance cannot be resolved by adjusting one's attitudes to match the (unethical) behavior, because doing so would imply a decline in moral standards and further threaten one's self-image (Barkan et al., 2015; Hochman et al., 2016).

Ethical dissonance can either be anticipated when contemplating an unethical action, such as lying (pre-violation), or experienced after an unethical action has occurred (post-violation). Acting as an internal gatekeeper, if the threat of anticipated dissonance is strong enough, it can deter unethical actions. However, people often use preemptive justifications by reframing unethical actions to fit within moral boundaries in a way that avoids the dissonance (Ayal, 2020; Shalvi et al., 2015). Examples include blurring ethical boundaries (e.g., claiming "there is no specific victim"), relying on negative social norms (e.g., "everyone does it"), or even justifying actions as altruistic when cheating benefits others (e.g., Hochman et al., 2021; Levine & Lupoli, 2022; Shuster et al., 2024; Weisel & Shalvi, 2015, 2022).

When dissonance is experienced following dishonest behavior, it poses a direct threat to people's moral identity (Aquino & Reed, 2002). In such cases, people can reduce dissonance and restore the paradoxical state of doing wrong while feeling moral (Gneezy et al., 2014; Shalvi et al., 2015). Frequently, corrections for experienced ethical dissonance involve moral-cleansing strategies aimed at righting the wrong and restoring the self-image (West & Zhong, 2015). Examples include donations, confessions, or prosocial activities. These atoning behaviors can lead to two patterns: moral bolstering, in which ethical behaviors continue, or moral licensing, in which the earned "moral credentials" justify future unethical behaviors, as if balancing a moral account (Effron & Helgason, 2023; Gneezy et al., 2014; Peer et al., 2014; Sachdeva et al., 2009). Finally, when ethical dissonance persists, people may distance themselves from their wrongdoing by adopting stricter moral standards and judging others harshly to maintain a virtuous public image (Berndsen & Gausel, 2015; Kennedy & Schweitzer, 2018).

3.4.1.1 Ethical dissonance and enforcement

Ethical dissonance is an internal cost of misconduct that is expected to occur independently of external detection. However, important interactions occur between internal ethical dissonance and the external detection and enforcement of disciplinary policies for misconduct.

First, the internal moral gatekeeper created by anticipated ethical dissonance can be combined with external detection to deter dishonesty. To strengthen this internal barrier, blocking the avenues of justification that reduce moral tension is essential (Ayal, 2020). Moral reminders, increased visibility, and fostering self-involvement and identification are all effective strategies for internal enforcement (Ayal et al., 2015; Peer et al., 2024). For example, Zickfeld et al. (2024) found honesty oaths—particularly those that specify unethical behavior and increase self-engagement—reduced dishonesty. As another example, a field study focused on train passengers in France showed posters with visibility cues (e.g., eye images) combined with social norm messages reduced fare evasion and lying (Ayal et al., 2021).

Second, increased ethical dissonance may make concealing guilt more difficult for offenders. Folklore suggests guilt is visible ("liar, liar, pants on fire"), and recent studies show lie detectors based on physiological arousal are more effective when ethical dissonance is active (e.g., lying for personal gain without justification) than when dissonance is dulled by altruistic justifications (Hochman et al., 2021).

An open question concerns a possible trade-off between ethical dissonance and detection. Specifically, external enforcement may reduce ethical dissonance because external incentives often undermine intrinsic motivation (Cerasoli et al., 2014; Deci, 1971; Ryan & Deci, 2000). This effect suggests that when detection is highly likely, ethical dissonance may be reduced. In addition, frequent detection may facilitate disengagement from internal moral values and replace them with practical and cynical criteria (Teodorescu, 2021). Similarly, punishment may lead to disengagement from the moral self ("I have paid my debt to society") rather than maintaining responsibility ("I will carry this guilt for the rest of my life") (Moore, 2015). Understanding the potential trade-off between internal ethical dissonance and external coercion could provide valuable insights.

3.4.1.2 Relational decisions and ethical dissonance

A critical factor in dishonest behavior is whether an individual acts alone or in concert with others. The involvement of others in the same unethical act can have opposing effects. On the one hand, groups can reduce an individual's ethical dissonance. Groups often provide a sense of anonymity and diffusion of responsibility, both of which can reduce the ethical dissonance (anticipated or experienced) associated with unethical behavior (e.g., Choo et al., 2019; Weisel & Shalvi, 2015). Another factor that contributes to reduced ethical dissonance is the group mechanism of persuasion. Exposure to others' rationalizations and alignment with negative norms provides additional justifications and reinforces the rationalizations one may already hold for unethical behavior (e.g., Janis, 2008; Bicchieri, 2005; Cialdini et al., 1991).

Conversely, groups may increase an individual's ethical dissonance. The presence of others can increase one's sense of visibility and self-awareness, potentially amplifying ethical dissonance and promoting peer monitoring (Cialdini & Goldstein, 2004; Duval & Wicklund, 1973; McKimmie, 2015).

The contrasting effects of group participation on an individual's ethical dissonance provide a valuable direction for future research. Studies could explore optimal "choice architectures" (Thaler & Sunstein, 2021) and identify strategies that promote the benefits of cooperation without incurring the costs of increased dishonesty (e.g., Bazerman & Tenbrunsel, 2011).

3.4.2 Attention

Eye-movement research has recently found dishonesty correlates with distinct patterns of attention and information processing (see Fiedler & Glöckner, 2015; Fosgaard et al., 2021; Hochman et al., 2016). One of these patterns is the "ethical blind spot"—a biased attentional tendency to focus on tempting and self-serving information at the expense of honest information (Bazerman & Tenbrunsel, 2011; Leib et al., 2019; Pittarello et al., 2015; Pittarello et al., 2016; Pittarello et al., 2019). To overlook what is right without feeling guilty (see Mazar et al., 2008), people need sufficient cognitive flexibility (Ayal et al., 2021; Hochman et al., 2021; Kunda, 1990; Peleg et al., 2021), which is facilitated by ambiguous situations where the distinction between right and wrong is unclear (see Hochman et al., 2016; Pittarello et al., 2015; Shalvi et al., 2011).

Building on our framework proposed in Fig. 2, blind spots have so far mainly been studied in situations where individuals act alone and in private settings. Thus, whether and to what extent blind spots would emerge in relational settings where the presence of others and perceived social norms guide behavior remains unclear (see Ayal et al., 2021; Lois & Wessa, 2021). Given that real-life ethical decisions are made in social contexts (Peleg et al., 2019), answering such a question is critical.

At the macro level, Dorrough et al. (2023) found more people offer bribes to individuals from countries perceived to be highly corrupt than from countries with lower perceived corruption. Gächter and Schulz (2016) showed countries with higher levels of corruption tend to have a higher prevalence of rule violations. Finally, a cross-cultural study by

Aycinena et al. (2022) found, perhaps counterintuitively, people cheated the most when the perceived social norm was that a small lie was just as unacceptable as a big lie. Given that norms shape people's moral perceptions (Ayal et al., 2021), future research should examine whether relational environments in which dishonesty is more prevalent provide individuals with sufficient cognitive flexibility to disregard ethical choices without compromising their self-concept.

At the micro level, Vives et al. found unethical behavior perpetrated by in-group members (but not by out-group members) increased immoral behavior. In addition, Ayal et al. (2021) found that a sign at the entrance of a train station in France stating that the majority pay for their train ticket reduced fare evasion by 50%. Clearly, the presence of others (either implicitly or explicitly), as well as the relational context that highlights people's interactions with each other, is important. But how would this affect blind spots? One possibility is that observing others breaking the rules would license people to cheat and increase blind spots. Another possibility is that it would make morality more salient (see Becker, 1968), which in turn may counteract blind spots and increase honesty (see Pittarello et al., 2019). Both alternatives can shed light on how the immediate environment and the nature of relationships between individuals shape attention toward moral norms.

Social relationships can also influence ethical behavior through social pressure and shared incentives (see Zanetti & Butera, 2023). However, whether people would ignore ethical considerations if they were pressured to cheat without personal gain is unclear. Thus, their attention to ethical norms may not align with their actual behavior, suggesting a disconnect between social influence and individual decision-making, a topic worthy of further investigation. In addition, shared interests within relationships may exacerbate ethical blind spots. Weisel and Shalvi (2015) found people engage in "corrupt collaboration" when their incentives align, compared with when they diverge. One can reasonably hypothesize ethical blind spots are more likely to emerge in situations where people's dishonest interests are aligned.

Finally, understanding the attentional process underlying moral behavior is crucial to our ability to detect and deter dishonesty. For example, Hochman et al. (2021) found people who cheated a lot to benefit others were less likely to be identified as cheaters in a lie-detection test than people who cheated a little to benefit themselves. These findings suggest good "reasons" may lead to guilt-free cheating behavior. As presented in our theoretical framework (see Fig. 2), much of this research has been conducted in "low" enforcement settings. An intriguing possibility would be to examine whether imposing fines of different sizes and frequencies for failing to report others' misbehavior would still lead people to fail to detect cues indicating dishonesty, even when such dishonesty is justified (e.g., prosocial lying).

In addition to practical implications, this research may shed light on how ethical blind spots may reduce our ability to detect deceptive behavior. Ultimately, this line of work has the potential to create effective interventions to correctly detect immoral acts by individuals.

3.4.3 Perspective

The common explanation for why people lie is that, while people want to achieve the more advantageous outcome, they also want to maintain their self-moral image. However, this explanation leaves open the question of what underlying cognitive process helps people maintain their self-image when they lie. More generally, one might ask how dishonest people view their behavior. Would a person who lied about a false outcome think of his or her action as "not lying" if that counterfactual outcome had been observed? Or perhaps this lie is a "little lie" —one that is within the norms? Of course, answering these questions requires addressing the question of what constitutes a lie and what is perceived as a lie.

Two lines of research have examined the question of what constitutes a lie. Although these two lines of research have produced consistent results, the underlying mechanisms are more complex than they appear at first glance. On the one hand, research by Bicchieri et al. (2022) suggests lying is more closely related to distorting beliefs about descriptive norms—what others actually do—than to revising prescriptive or ethical norms about what should be done. Their findings suggest believing that many people lie does not necessarily mean one now views lying as ethical. Instead, people may engage in dishonest acts based on a biased perception that such behavior is widespread and conforms to a misperceived descriptive norm, even while their beliefs about moral standards remain intact. This finding suggests the underlying psychological mechanism of dishonesty stems more from conforming to perceived social norms than from actively reshaping one's ethical beliefs.

On the other hand, research by Mazar et al. (2008) on self-concept maintenance theory posits that people use moral flexibility and rationalization to allow dishonest behavior that violates ethical norms, but only to the extent that it still allows them to maintain a positive self-concept as an honest person. The underlying process is to adjust one's ethical standards through motivated reasoning just enough to rationalize some unethical acts as morally permissible, while avoiding a complete shattering of one's moral self-image. Thus, self-deception about the ethicality of the behavior itself, rather than just perceptions of the descriptive norm, enables people to cheat within reasonable limits.

Several findings can be interpreted as supporting the latter account. Schurr et al. (2012) examined the effects of different choice procedures on dishonest behavior. In one of their experiments, participants played a 20-question trivia game in which they were shown the correct answers before being asked to report whether they knew the answers, while the experimenters had no way of detecting whether a particular participant was lying. Participants earned money each time they claimed to have known the correct answer, with higher payouts for difficult questions. The experiment manipulated the perspective of choice: participants in the narrow-perspective condition chose between an easy and a difficult question before each trial, whereas those in the broad-perspective condition decided in advance how many easy and difficult questions they wanted to answer in the 20 trials. Consistent with the self-concept maintenance account, the study found that in the narrow-perspective condition, when faced with a sequence of 20 repeated continuous choices between easy and difficult questions, participants were more likely to choose difficult (more profitable) questions and more likely to report knowing the correct answers than in the broad-perspective condition. Thus, despite the identical nature of the task in both conditions, the higher rate of dishonesty in the narrow-perspective condition suggests participants did not change their ethical norms, but rather used a flexible perception of lying to justify their behavior. Rilke et al. have reported similar findings.

3.4.4 Honesty as an automatic versus controlled behavior

Several lines of research have proposed cognitive control is crucial in resolving this conflict between serving self-interest and maintaining an honest self-concept (Abe & Greene, 2014; Greene & Paxton, 2009; Mead et al., 2009; Verschuere et al., 2018). However, the precise nature of how cognitive control resolves this conflict remains controversial. Scientists and philosophers have long debated the role of cognitive control in (dis)honest decision-making: according to the grace hypothesis, people naturally default to honesty. According to the will hypothesis, people naturally default to self-interest. Honesty requires willpower, so people must exercise cognitive control to override their dishonest default.

In support of the will hypothesis, some studies have reported that impairing participants' capacity for cognitive control, such as when they are exhausted by demanding tasks, sleep deprived, anxious, or under time pressure, makes them less likely to be honest (Barnes et al., 2011; Gunia et al., 2012; Shalvi et al., 2012; Zhang et al., 2020). However, caution should be exercised in interpreting these findings, because a substantial part of this body of evidence relies on now highly controversial experimental manipulations such as ego depletion (Vohs et al., 2021), and replicating study findings with uncontroversial manipulations appears to be difficult (van der Cruyssen et al., 2019). A number of fMRI studies show that when regions in the cognitive control network, such as the dorsolateral and ventrolateral prefrontal cortex, are activated, dishonest individuals can suppress selfish impulses and avoid lying (Abe & Greene, 2014; Abe et al., 2018; Dogan et al., 2016; Knoch et al., 2006; Yin et al., 2016). Early work also suggests that when activity in these areas is disrupted by transcranial magnetic stimulation (TMS) or lesions, selfish behavior predominates (Knoch et al., 2006; Zhu et al., 2014). However, caution is warranted in interpreting TMS results, because a meta-analysis revealed they are inconsistent and effect sizes are small (see Ganis, 2014, 2015, for a critical review).

Evidence in favor of the grace hypothesis comes from studies showing people respond more quickly when asked to tell the truth than when asked to lie (see the meta-analysis of 114 studies with n = 3307; Suchotzki et al., 2017; Verschuere et al., 2018). Moreover, the fact that most people report being truthful most of the time (Serota et al., 2010; Debey et al., 2015; Halevy et al. 2014) is also consistent with the "truth default"—the idea that telling the truth is the dominant and automated behavior in human communication (Levine, 2014). That the truth would come naturally also explains the increased activity in cognitive control regions of the brain when lying (for a meta-analysis, see Christ et al., 2009). Further support for the grace hypothesis comes from developmental studies showing children begin life completely honest and only gradually learn to lie as their cognitive control capacities develop and they acquire the ability to inhibit (honest) behavior (Talwar & Lee, 2002, 2008). Moreover, introducing time pressure in contexts where decision-makers do not have immediate access to their payoff-maximizing strategy makes them more likely to resort to telling the truth. This finding suggests the true state of the world serves as an anchor to guide individuals when they are unable to fully calculate the consequences of their available actions (Capraro et al., 2019; Capraro, 2017; Lohse et al., 2018).

3.4.4.1 The moral default

The will and grace hypotheses seem to be in direct conflict: they cannot both be correct. Several explanations have been proposed, including the incentives and self-interest that deception brings (Verschuere & Shalvi, 2014), the social consequences of lying (Köbis et al., 2019), and the accessibility of the payoff-maximizing strategy (Capraro, 2024). Another factor that may help reconcile these two hypotheses is the recognition of significant individual differences in (dis)honesty (Speer et al. 2020, 2021, 2022b; Yin & Weber, 2019). Several personality traits, such as psychopathic traits, impulsivity, creativity, greed, and implicit attitudes toward (dis)honesty, are associated with these variations in unethical behavior and the tendency to cheat (Anderman et al., 2009; Halevy et al., 2014; Hatta et al., 2022; Seuntjens et al., 2019; Speer et al., 2022b).

Recent studies have investigated whether these dispositions toward (dis) honesty are encoded in stable patterns of the resting brain. They used resting-state fMRI, which measures fluctuations in connectivity patterns when people are not engaged in a specific task. Variability in resting-state connectivity has been found to be substantial across individuals, but stable within an individual over time (Cao et al., 2014; Finn et al., 2015; Zuo & Xing, 2014). Metaphorically, resting-state connectivity patterns could be compared to a brain "fingerprint." It has been associated with individual differences in personality as well as social decision-making (Cai et al., 2020; Li et al., 2013; Nostro et al., 2018). Resting-state studies of (dis)honesty have found higher connectivity between self-referential thinking and the reward network is associated with more honesty. This finding has been replicated across age groups (Yin et al., 2021) and cultures (Pang et al., 2021), and a recent study has also linked similar connectivity patterns to driving violations (Ju, 2023).

Taken together, compelling evidence shows considerable heterogeneity in (dis)honesty, and emerging evidence suggests this heterogeneity is encoded in stable patterns of connectivity in individuals' brains. Thus, participants may be distributed along a continuum: whereas some—arguably most—people are more predisposed to honesty, others—probably a minority—are more predisposed to cheating (meta-analyses suggest this distribution appears to be skewed towards honesty: Abeler et al., 2019; Gerlach et al., 2019). On the one hand, participants have a default bias toward honesty, which is associated with more self-referential thinking when faced with the opportunity to cheat. On the other hand, people have a strong predisposition toward dishonesty, and their decisions are driven by reward anticipation (Speer et al., 2022a). The predisposition toward honesty or (dis)honesty has been termed "moral default" (Speer et al., 2022a). Thus, whether situational demands (e.g., stress, demanding tasks, or sleep deprivation) make people more or less likely to lie depends on their moral default. Those who are honest would be more honest in the presence (vs. absence) of pressure, whereas those who are dishonest would be more dishonest in the presence (vs. absence) of pressure. Now, this idea of a moral default raises several important questions (see Table 1).

Research in this area could greatly benefit from a measure of moral default that is easy to administer. This approach would allow such a measure to be routinely included in research, extending findings from cognitive neuroscience to psychology, communication science, and behavioral economics. Beginning this quest with established measures, such as self-report measures of psychopathy (Muris et al., 2017) or honesty (Thielmann et al., 2017), seems prudent. The key research question is whether this moral-standard measure reliably explains substantial heterogeneity in self-report, behavioral, and neuroimaging research findings. Does a simpler alternative exist to resting-state connectivity to reliably assess moral default and explain heterogeneity in self-report, behavioral, and neural responses to (dis)honesty? Do current findings, typically obtained by testing individuals in isolation, extend to real-time, dynamic, two-way interactions? And does science on the intuitive nature of (dis)honesty allow for the development of new ways to distinguish lies from truth? Do new insights into the intuitive nature of (dis)honesty allow us to design interventions to promote greater honesty in society?

Such a moral-default measure allows us to examine how people with different moral defaults resolve the moral conflict associated with lying. Research suggests conflict is often resolved through compromise, where people behave dishonestly enough to benefit, but honestly enough to maintain a positive self-concept. Recent neuroimaging studies (Speer et al. 2020, 2021) examined how cognitive control balances rewards and self-image. The studies found dishonest people needed cognitive control to be honest. Conversely, honest participants used cognitive control to cheat. Thus, the role of cognitive control varies with an individual's moral standard, with honest individuals using it to suppress honesty at times, whereas cheaters use it to act honestly at times.

Current research on dishonesty tends to study individuals in isolation—a respondent to a survey, a player in a behavioral economics game, or a person lying in an MRI scanner. But deception is inherently a social behavior. So where is the other? Recently, behavioral economists have changed the games they play, so they can now study interactions between individuals (Leib et al., 2021). Similarly, neuroimaging research on (dis)honest decision-making has primarily used paradigms that do not allow for real-time, dynamic, two-way interaction. A promising future avenue may be to move away from these conventional paradigms by studying spontaneous and voluntary dishonesty and its neurocognitive underpinnings in naturalistic, real-time interactions such as conflictual conversations. Hyperscanning can be used for this purpose. In hyperscanning, neuroimaging data are collected from dyads engaged in two-way interactions, dynamically shaping the conversation while their neural activity is recorded in real time. The advantage of this approach is that it allows researchers to capture the emergent properties of these interactions. In the context of dishonesty, it allows us to assess (a) what neurocognitive underpinnings precede the decision to cheat in a naturalistic setting, (b) how the interaction partner perceives this dishonest act, and (c) how dishonesty affects the interaction and the underlying neural dynamics between the two interaction partners.

3.4.4.2 The applied agenda: Detecting deception

Although several basic research questions remain to be answered, considering possible applications of more fundamental deception research is worthwhile. Here, we consider one such application, namely, deception detection.

About a decade ago, Vrij et al. (2008) proposed a novel approach to deception detection rooted in the current understanding of deception. That people, on average, take longer to lie than to tell the truth (Suchotzki et al., 2017) supports the view that truth-telling is more automatic and that lying requires cognitive control. The novel idea was to design interventions that further increase the cognitive load of lying so that it becomes apparent and can be used as a cue to lie. To impose cognitive load, the interviewer could ask unexpected questions, tell the story in reverse order, or have the interviewee perform an additional task. Similarly, Speer et al. (2023) found exposing people to stress reduces cognitive control and reveals moral deficiencies (Speer et al., 2023). Whereas some studies have been successful, others have failed to replicate these findings (Brimbal et al., 2023), and a meta-analysis showed imposing cognitive load does not readily improve deception detection (Mac Giolla & Luke, 2021; Verschuere et al., 2018). Again, however, an unexplored area of research is whether individual differences in the propensity to lie or tell the truth might partly explain the divergent research findings.

Are people more likely to be honest or dishonest? Research from several disciplines seems to converge on the idea that substantial individual differences exist. For the vast majority, the moral default is to be honest, but a minority may tend to be dishonest. Research would benefit from a simple measure of the moral default, not only to study how cognitive conflict in moral decision-making is resolved, but also to tailor interventions to promote honesty and perhaps improve deception detection techniques.

3.4.5 Embodied (dis)honesty

Embodied cognition, an approach that has profoundly influenced the neuropsychological study of mental processes over the past two decades, postulates that even low-level, body-related exteroceptive and interoceptive signals mediate high-level, complex processes (Barsalou, 2008). Body and brain signatures of "embodiment" have been provided for motor-cognitive (Moro et al., 2008; Urgesi et al. 2004, 2007a, 2007b), emotional (de Gelder, 2006), highly abstract (e.g., grammar) (Liuzza et al., 2011; Williams et al., 2009), and even ineffable functions such as self-transcendence (Urgesi et al., 2010), sense of fairness (Lenggenhager et al., 2013), and observation of immoral acts (Liuzza et al., 2015). The bodily dimension of morality is also suggested by the need to wash one's body after an immoral act (Lee & Schwarz, 2010), the increased severity of moral judgments associated with bodily dirtiness (Schnall et al., 2008)—an effect that, however, was not confirmed in a direct replication study (Johnson et al., 2014)—and the effect of physical cleanliness in reducing the distressing consequences of unethical behavior and threats to one's moral self-image (Zhong & Liljenquist, 2006).

As noted above, (dis)honesty involves a conflict between the temptation to gain some benefit and the desire to conform to personal and societal moral norms or to avoid aversive social consequences (Mazar et al., 2008; Mead et al., 2009). It is also influenced by complex variables related to the individual, such as personality (Johnson et al., 2005; Kashy & DePaulo, 1996; Panasiti et al., 2011), creativity (Vincent & Kouchaki, 2016), emotional and cognitive intelligence (Pittarello et al., 2017; Sarzyńska et al., 2017), and affiliation (e.g., religious, Arbel et al., 2014; Bar-El & Tobol, 2017) with a social group. Exploring whether individual dishonesty is also influenced by the level of unethical behavior prevalent in a given community has not yielded clear results, with data supporting (Gächter & Schulz, 2016) or contradicting (Mann et al., 2016) the primary role of societal habits in promoting individual dishonesty. Interestingly, however, the notion that people may be fundamentally similar across societies in their tendency to deceive others does not imply dishonest behavior is independent of context but calls for studies of the relationships between dishonesty and body-related signals in the brain of people who act unethically, individually or as a group.

A research program on embodied morality has been developed around the notion of bodily self-consciousness (BSC), that is, a basic primitive of the "self" (Berlucchi & Aglioti, 1997; 2010), which consists mainly of the sense of ownership of one's own body, that is, the sense of ownership (Blanke & Metzinger, 2009; Blanke, 2012), and of being in control of one's own actions, that is, the sense of agency, a mental property inextricably linked to the notion of free will and responsibility (Haggard, 2017). BSC arises from the integration of body-related exteroceptive (e.g., visual and auditory) and interoceptive (e.g., cardiac, respiratory, bowel motility, hunger, thirst, temperature, proprioception, pain, itch, and sensory touch) signals. Importantly, the BSC is critical not only for defining the sense of self, but also for distinguishing the self from others (Berlucchi &Aglioti, 2010; Salomon, 2017), which is a central process in social cognition (Decety & Sommerville, 2003). The strength of this self-other distinction can be deeply influenced by social dimensions, as evidenced by situations in which individuals feel a sense of fusion (shared BSC) with their in-groups, both in terms of ownership (e.g., feeling like the same entity) and agency (e.g., feeling like acting as a single entity; Besta et al., 2016; Swann et al., 2012). This sense of fused identity supports extreme pro-group behavior (Swann & Buhrmester, 2015), which, although considered moral by some, can be devastating for many (e.g., suicidal terrorist attacks in the name of group identity; Swann et al., 2009, 2010). In this sense, changes in the BSC may influence individual dishonesty. Several hypotheses related to this issue have been tested using the Temptation to Lie Card Game (TLCG) as a proxy for morality (Panasiti et al., 2011). The TLCG is an experimental task in which participants spontaneously decide to lie (egoistically or altruistically) to another player (typically a research confederate) who must try to pick the trump card (i.e., the ace of hearts) from two covered cards. Favorable trials are those in which the chosen card is a loss to the participant, and vice versa in unfavorable trials. Importantly, the opponent cannot see directly whether he has won or lost, but receives this information from the participant, who can decide to lie and thus reverse the outcome of the game (i.e., win if he has lost [self-gain lie]–or lose if he has won [other-gain lie]), or to tell the truth. Because winning involves taking money out of a common pocket, self-gain lies are tantamount to stealing money from another individual, thus committing specific moral violations. The risk of being caught in a deception and thus jeopardizing one's reputation varied across conditions. The moral personality of the participants was also measured.

The TLCG proved to be a suitable tool for studying social deception under well-controlled and still rather ecological conditions. Indeed, the number of self-gain lies was influenced by reputation risk, modulated by moral-related personality traits such as Machiavellianism versus social reward dependence, and modulated by the perceived warmth or competence of the interactive partners (Azevedo et al., 2018; Panasiti et al., 2014, 2016). Moreover, bodily or brain signatures of the process of self-gain have been provided (Dupont et al., 2023; Panasiti et al. 2014, 2016). In addition to demonstrating a link between the BSC and moral personality (Scattolin et al., 2022a), using the TLCG as a proxy for morality allowed us to show increasing or decreasing the sense of ownership over an artificial virtual body decreases and increases individual honesty (Scattolin et al., 2022b). Similarly, a study of people with body-integrity disorders shows deceptive choices expressed with a disowned body part reduce body ownership (Scattolin et al., 2023). Also note that when reputation was at risk (i.e., the other player knew they were lying), participants told significantly fewer egoistic lies than when their decisions were secret, an effect that was moderated by the internal bodily signals associated with cardiac interoception (Vabba et al., 2022). Finally, mindfulness meditation may reduce dishonest behavior through increased regulation of interoceptive awareness (Feruglio et al., 2024). This finding supports the notion of a close connection between the body and honesty and paves the way for moral self-training programs.

3.5 Situational factors

3.5.1 Honesty pledges

Honesty oaths and pledges have been used by philosophers (Plato's Apology of Socrates), physicians (the Hippocratic Oath), religious leaders (the Papal Oath), judges (the British Judicial Oath), armies (the Oath of Enlistment), and industry (the Dutch Bankers' Oath). Honesty pledges are also a ubiquitous behavioral intervention to reduce dishonesty (Hertwig & Mazar, 2022). Here, we focus on the current state of knowledge about the effectiveness of honesty pledges, their potential applications, and future directions for their research.

Several studies have examined honesty pledges and found significant effects on reducing dishonesty (e.g. Babin et al., 2022; Heinicke et al., 2019; Jacquemet et al., 2019; le Maux & Necker, 2023; Peer et al., 2024; Zickfeld et al., 2023). For example, Peer and Feldman (2021) showed honesty promises made ex ante (before the opportunity to lie) can significantly and consistently reduce dishonesty, even when accompanied by sanctions. These findings are important, especially in light of reported replication failures (Kristal et al., 2020) and concerns about data credibility. Importantly, a meta-analysis found commitment to honesty norms through oaths or pledges can increase honest behavior (Zickfeld et al., 2024). The study examined several situational and individual factors that influence dishonest behavior, such as punishment, age, gender, and financial rewards. For example, the authors found task repetition and public (vs. private) decision-making increased honesty.

More recently, a mega-study examined the effect of 21 different honesty oaths, varying in theoretical basis, placement, and type/involvement, on honesty in a tax-evasion task (Zickfeld et al., 2024). The authors report that 10 of these oaths increased compliance (relative to the control), with the most successful being those aimed at reducing ambiguity (Skowronek, 2022). Similarly, another recent study highlighted the importance of designing the pledge to require high engagement (e.g., by asking participants to retype the text), which increases attention to the text of the pledge and its effectiveness (Peer et al., 2024). In terms of the framework in Fig. 2, most studies of pledges appear to have focused on solitary decision-making, and most of them have evaluated situations where enforcement was nil or low.

Field studies of honesty pledges are rarer, and their results are mixed. The Social and Behavioral Science Team in the US (2015) reported on an experiment that promoted more accurate self-reporting of sales, resulting in more accurate collection of fees from vendors of goods and services to the federal government. Adding an honesty pledge at the top of the online data-entry form resulted in vendors self-reporting significantly more sales than without the honesty pledge, contributing to an additional \$1.59 million in fees collected by the federal government within a single quarter. However, a field study in Guatemala found no effect of the pledge on tax reporting (Kettle et al., 2017). Similarly, a field study using a no-cheating pledge before an exam also found no effect on student honesty (Cagala et al., 2024).

For honesty pledges to have impactful applications, more field research is needed. Potential applications that could be tested through field research include many government-citizen interactions, such as increasing the accuracy of tax reporting and ensuring honest claims of eligibility for benefits or public resources. Honesty pledges could also replace legal affidavits to reduce administrative burdens on individuals and businesses and speed up business licenses without increasing public risk. More field research is needed for multiple reasons: to understand whether the experimental findings on honesty pledges generalize to real-world behavior; and to understand how honesty pledges should be designed for different situations, such as those with very high stakes, strong competition, and conflicting social norms. Such field research and applications may also allow us to understand whether or when honesty pledges can be used not only to reduce dishonesty, but also as tools to promote mutual trust between governments and their citizens, between managers and their employees, and among individuals who wish to live in a trustworthy society.

Several research questions remain to be addressed in relation to honesty pledges and their effectiveness in reducing dishonest behavior: (a) To what extent can the effects of honesty pledges observed in online or laboratory settings be generalized to real-life situations, and what are the barriers and effectiveness of implementing honesty pledges in different real-life contexts, such as tax reporting, public service, or business ethics? (b) What specific elements of honesty pledges (e.g., wording, format, timing) are most effective in reducing dishonest behavior? (c) How do different situational factors (e.g., punishment severity, task repetition, public vs. private decision making) influence the effectiveness of honesty pledges? (d) How can honesty pledges be integrated into existing legal and administrative systems to improve compliance? (e) Do the effects of honesty pledges persist or fade over time, and how often should honesty pledges be renewed or reinforced to maintain their effectiveness? (f) What are the psychological mechanisms underlying the effectiveness of honesty pledges? (g) Can honesty pledges foster mutual trust between governments and citizens, managers and employees, or within communities, and what are the potential unintended consequences of honesty pledges, such as increased skepticism or resistance if perceived as coercive?

3.5.2 Artificial intelligence

Artificial intelligence (AI), systems that use statistical models to analyze their environment and take action—with a degree of autonomy—to achieve specific goals (European Commission, 2011), is impacting our lives. For example, AI tools help human resources departments screen potential employees' resumes, set prices in online markets, and advise individuals on purchasing and investment decisions. Furthermore, AI is increasingly taking on social roles that were once the domain of humans, due to the advancement of large language models (LLMs), which brings immense potential but also new risks (Starke et al., 2024). Indeed, tools such as ChatGPT and Replika ("The AI companion who cares") are becoming advisors, friends, and collaborators. With the rise of AI, a growing body of work has begun to examine how AI tools shape honesty (Köbis et al., 2021). But as with any technology, AI can be used for good or evil.

On the one hand, when facing an ethical dilemma, individuals can assign blame not only to fellow humans but also to AI agents (Hohenstein & Jung, 2020). This ability, in turn, frees people to engage in dishonest acts while alleviating guilt and maintaining a positive self-concept. For example, recent work has shown people are willing to follow dishonesty-promoting AI advice and forgo their honesty for financial gain, even when they know AI generated the advice (Leib et al., 2024). Similarly, people use AI agents as delegates to do their "dirty work" for them, assigning morally questionable tasks such as coordinating prices in online markets (Calvano et al., 2020), negotiating on their behalf (Bianchi et al., 2024), and lying for financial gain (Köbis et al., 2024).

On the other hand, AI can also be used as a tool to enhance ethical behavior, for example, by reducing corruption. AI is particularly useful because it can sift through large datasets, detect suspicious patterns, and "blow the whistle" without fear of repercussions (Köbis et al., 2022; Starke et al., 2023). In a laboratory setting, AI can already detect lies more accurately than the average participant (von Schenk et al., 2024). Extrapolating these findings outside of the lab, such technology could eventually dissuade people from (blatant) lying. Given the framework of the manuscript (see Fig. 2), we next consider how AI can shape honesty in relational, hybrid team settings, and in settings where detection and enforcement are possible.

3.5.2.1 AI in hybrid team settings

In relational, hybrid team settings, AI can take on different roles: it can be a principal managing groups (Koster et al., 2022) or a team member and collaborator (Shirado et al., 2023). Teams composed of only humans are, on average, more dishonest than individuals working alone (Leib et al., 2021; Weisel & Shalvi, 2015). By contrast, teams composed of humans and AI members can be either more or less honest, depending on how the AI is programmed. However, research on this topic is limited. More work is needed to understand how AI as a team member influences group (dis) honesty, and how different group compositions shape it. Specifically, how do different proportions of AI agents influence a group of humans? And is the effect symmetric when comparing honest and dishonest AI agents?

As a managing principal, AI can potentially form groups and influence their dynamics to promote honesty. When groups are composed of rule followers, they are more honest than groups composed of rule breakers (Gross & de Dreu, 2021). AI could play an important role in this process by, for example, automatically classifying individuals as rule followers or rule breakers and optimally assembling teams based on this classification. However, an important challenge is that such AI systems are currently unable to reliably detect lies in very complex contexts, often due to a lack of sufficient high-quality training data (Boffey, 2018).

3.5.2.2 AI in settings with detection and enforcement

Introducing the possibility of detection and enforcement into environments where people can lie increases the fear of punishment and subsequently decreases lying (Teodorescu et al., 2021). When detection is possible, AI can shape honesty either by being a tool through which people lie or by being a tool through which lies are detected and honesty is enforced. When used as a tool through which people lie, AI can be used as a scapegoat, sharing some of the blame and subsequent punishment with dishonest humans. Current work on the extent to which blame is shifted to AI is mixed. Some work shows that compared with machines, humans are perceived as more blameworthy and responsible for immoral decisions (Wilson et al., 2022) or mistakes with moral consequences (Awad et al., 2019). Conversely, other research suggests AI agents and humans are perceived as equally responsible for wrongdoing (Lima et al., 2021). Systematically understanding when and why blame is shifted to AI is crucial, both theoretically and practically. As AI-human interactions become a larger part of our lives, mistakes and unethical acts involving both parties will occur. Societal judgments about such cases can help inform the legal framework for addressing human corruption facilitated by AI.

Considering AI as a detection and potential enforcement tool, AI can process vast amounts of data to identify anomalies and patterns that indicate fraud or misconduct (Deodhar et al., 2022). For example, AI systems in financial institutions can detect unusual transaction patterns that may indicate fraudulent activity (Xu et al., 2024). However, the perception of AI as a legitimate and fair enforcer of honesty is critical. Research is needed to examine how the presence of AI detection tools affects the ethical behavior of individuals, and when AI tools are perceived as legitimate in such detection roles. For example, privacy concerns and fears of creating a "big brother" environment may undermine the perceived legitimacy of using AI in this role. However, studying fraudulent behavior under various enforcement policies, as proposed in the current framework (see Fig. 2), can guide the development of interventions that leverage technological advances without the need to collect personally identifiable information, thereby mitigating privacy concerns (e.g., Erev et al., 2024). Addressing concerns about privacy, transparency, accountability, and potential bias in AI algorithms is essential to ensuring these tools are both effective and trusted by the people they monitor.

4. Concluding remarks: The future of the science of honesty

The study of honesty has provided invaluable insights into human behavior, especially in contexts where no external consequences exist and individuals act alone. However, to truly understand the complexities of honesty, future research must move beyond controlled settings to explore how honesty operates in more dynamic, real-world environments.

Our framework (Fig. 2) suggests a path forward by highlighting the need to investigate underexplored areas, such as settings with external enforcement and relational decision-making contexts. Through a comprehensive review of the state of the art in various disciplines, we show how broadening the scope of research addresses important theoretical questions, as summarized in Table 1.

A cross-disciplinary approach is essential for a deeper understanding of honesty. By integrating perspectives from psychology, economics, sociology, and computer science (with an emphasis on AI), we can achieve a more nuanced and realistic view of honesty. This approach not only enriches theoretical knowledge, but also offers practical insights for promoting honesty in all domains, from personal relationships to professional settings.

In conclusion, although significant progress has been made in the science of honesty, much remains to be explored. We encourage researchers to advance this agenda by examining less understood but crucial contexts to deepen our understanding of one of humanity's most cherished virtues. Through such efforts, we can better appreciate the complex interplay between truth, deception, and the many factors that shape human behavior.

Acknowledgments

Except for the first author, author order was randomly determined. Shaul Shalvi was supported by the HORIZON EUROPE European Research Council (ERC-CoG-865931) and the Dutch Research Council (Nederlandse Organisatie voor Wetenschappelijk

Onderzoek; Vi.Vidi.195.137) grants. Van Rooij's contribution was supported through grants from the European Research Council (ERC-2018-CoG-HomoJuridicus-817680) and the University of Amsterdam Research Priority Area Organizational Ethics.

Dr. M.E. Kret was supported by an ERC Starting grant (804582) and an NWO 016.Vidi.185.036. Jayawickreme and Furr were supported by grant #61842 from the John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation. Yuval Feldamn was supported by ERC Grant number: 101054656/Project acronym: VCOMP. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 771391)

References

Abe, N., & Greene, J. D. (2014). Response to anticipated reward in the nucleus accumbent predicts behavior in an independent test of honesty. *Journal of Neuroscience, 34*(32), 10564–10572. https://doi.org/10.1523/JNEUROSCI.0217-14.2014.

Abi-Esber, N., Abel, J. E., Schroeder, J., & Gino, F. (2022). "Just letting you know…" Underestimating others' desire for constructive feedback. *Journal of Personality and Social Psychology, 123*(6), 1362.

Abe, N., Greene, J. D., & Kiehl, K. A. (2018). Reduced engagement of the anterior cingulate cortex in the dishonest decision-making of incarcerated psychopaths. *Social Cognitive and Affective Neuroscience, 13*(8), 797–807. https://doi.org/10.1093/scan/nsy050.

Abeler, J., Nosenzo, D., & Raymond, C. (2019). Preferences for truth-telling. *Econometrica: Journal of the Econometric Society, 87*(4), 1115–1153. https://doi.org/10.3982/ECTA14673.

Alfonso-Costillo, A., Brañas-Garza, P., & López-Martín, M. C. (2022). Does the die-under-the-cup device exaggerate cheating? *Economics Letters, 214*, 110424. https://doi.org/10.1016/j.econlet.2022.110424.

Anderman, E. M., Cupp, P. K., & Lane, D. (2009). Impulsivity and academic cheating. *The Journal of Experimental Education, 78*(1), 135–150. https://doi.org/10.1080/00220970903224636.

Anderson, L. R., & Stafford, S. L. (2003). An experimental analysis of rent seeking under varying competitive conditions. *Public Choice, 115*(1), 199–216.

Apel, R. (2013). Sanctions, perceptions, and crime: Implications for criminal deterrence. *Journal of Quantitative Criminology, 29*(1), 67–101. https://doi.org/10.1007/s10940-012-9170-1.

Aquino, K., & Reed, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology, 83*(6), 1423–1440. https://doi.org/10.1037/0022-3514.83.6.1423.

Arbel, Y., Bar-El, R., Siniver, E., & Tobol, Y. (2014). Roll a die and tell a lie – what affects honesty? *Journal of Economic Behavior & Organization, 107*, 153–172. https://doi.org/10.1016/j.jebo.2014.08.009.

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review, 11*(2), 150–166. https://doi.org/10.1177/1088868306294907.

Awad, E., Levine, S., Kleiman-Weiner, M., Dsouza, S., Tenenbaum, J. B., Shariff, A., ... Rahwan, I. (2019). Drivers are blamed more than their automated cars when both make mistakes. *Nature Human Behaviour, 4*(2), 134–143. https://doi.org/10.1038/s41562-019-0762-8.

Ay, F. C., Berge, J. W., & Nødtvedt, K. B. (2024). Strategic curiosity: An experimental study of curiosity and dishonesty. *Journal of Economic Behavior & Organization, 217*, 287–297.

Ayal, S. (2020). Comments on "*The Law of Good People*": Regulating justified dishonesty. *Jerusalem Review of Legal Studies, 22*(1), 38–48. https://doi.org/10.1093/jrls/jlz013.

Ayal, S., Celse, J., & Hochman, G. (2021). Crafting messages to fight dishonesty: A field investigation of the effects of social norms and watching eye cues on fare evasion. *Organizational Behavior and Human Decision Processes, 166*, 9–19. https://doi.org/10.1016/j.obhdp.2019.10.003.
Ayal, S., Gino, F., Barkan, R., & Ariely, D. (2015). Three principles to REVISE people's unethical behavior. *Perspectives on Psychological Science, 10*(6), 738–741. https://doi.org/10.1177/1745691615598512.
Aycinena, D., Rentschler, L., Beranek, B., & Schulz, J. F. (2022). Social norms and dishonesty across societies. *Proceedings of the National Academy of Sciences, 119*(31), https://doi.org/10.1073/pnas.2120138119.
Azevedo, R. T., Panasiti, M. S., Maglio, R., & Aglioti, S. M. (2018). Perceived warmth and competence of others shape voluntary deceptive behaviour in a morally relevant setting. *British Journal of Psychology, 109*(1), 25–44. https://doi.org/10.1111/bjop.12245.
Babin, J. J., Chauhan, H. S., & Liu, F. (2022). You can't hide your lying eyes: Honesty oaths and misrepresentation. *Journal of Behavioral and Experimental Economics, 98*, 101880. https://doi.org/10.1016/j.socec.2022.101880.
Babin, J. J., & Chauhan, H. S. (2024). Replication: Erat and Gneezy's white lies paradigm. *Journal of Economic Psychology, 102*, 102704. https://doi.org/10.1016/j.joep.2024.102704.
Bandura, A. (2002). Selective moral disengagement in the exercise of moral agency. *Journal of Moral Education, 31*(2), 101–119. https://doi.org/10.1080/0305724022014322.
Bar-El, R., & Tobol, Y. (2017). Honesty toward the holy day. *Journal of Behavioral Experimental Economics, 68*, 13–17. https://doi.org/10.1016/j.socec.2017.02.005.
Barkan, R., Ayal, S., & Ariely, D. (2015). Ethical dissonance, justifications, and moral behavior. *Current Opinion in Psychology, 6*, 157–161. https://doi.org/10.1016/j.copsyc.2015.08.001.
Barnes, C. M., Schaubroeck, J., Huth, M., & Ghumman, S. (2011). Lack of sleep and unethical conduct. *Organizational Behavior and Human Decision Processes, 115*(2), 169–180. https://doi.org/10.1016/j.obhdp.2011.01.009.
Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639.
Baum, J., Rabovsky, M., Rose, S. B., & Abdel Rahman, R. (2020). Clear judgments based on unclear evidence: Person evaluation is strongly influenced by untrustworthy gossip. *Emotion, 20*(2), 248–260. https://doi.org/10.1037/emo0000545.
Bazerman, M. H., & Tenbrunsel, A. E. (2011). *Blind spots*. Princeton University Press. https://doi.org/10.1515/9781400837991.
Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy, 76*(2), 169–217. http://www.jstor.org/stable/1830482.
Beersma, B., & van Kleef, G. A. (2011). How the Grapevine keeps you in line. *Social Psychological and Personality Science, 2*(6), 642–649. https://doi.org/10.1177/1948550611405073.
Berlucchi, G., & Aglioti, S. (1997). The body in the brain: Neural bases of corporeal awareness. *Trends in Neurosciences, 20*, 560–564. https://doi.org/10.1016/s0166-2236(97)01136-3.
Berlucchi, G., & Aglioti, S. M. (2010). The body in the brain revisited. *Experimental Brain Research, 200*, 25–35. https://doi.org/10.1007/s00221-009-1970-7.
Berndsen, M., & Gausel, N. (2015). When majority members exclude ethnic minorities: The impact of shame on the desire to object to immoral acts. *European Journal of Social Psychology, 45*(6), 728–741. https://doi.org/10.1002/ejsp.2127.
Besta, T., Mattingly, B., & Błażek, M. (2016). When membership gives strength to act: Inclusion of the group into the self and feeling of personal agency. *Journal of Social Psychology, 156*, 56–73. https://doi.org/10.1080/00224545.2015.1053838.

Bianchi, F., Chia, P.J., Yuksekgonul, M., Tagliabue, J., Jurafsky, D., & Zou, J. (2024). How well can llms negotiate? negotiationarena platform and analysis. arXiv preprint arXiv:2402.05863.
Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.
Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press. https://doi.org/10.1017/CBO9780511616037.
Bicchieri, C., & Chavez, A. (2010). Behaving as expected: Public information and fairness norms. *Journal of Behavioral Decision Making, 23*(2), 161–178. https://doi.org/10.1002/bdm.648.
Bicchieri, C., Dimant, E., & Sonderegger, S. (2023). It's not a lie if you believe the norm does not apply: Conditional norm-following and belief distortion. *Games and Economic Behavior, 138*, 321–354. https://doi.org/10.1016/j.geb.2023.01.005.
Bicchieri, C., & Dimant, E. (2019). Nudging with care: The risks and benefits of social information. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3319088.
Bicchieri, C., Dimant, E., & Xiao, E. (2021). Deviant or wrong? The effects of norm information on the efficacy of punishment. *Journal of Economic Behavior & Organization, 188*, 209–235. https://doi.org/10.1016/j.jebo.2021.04.002.
Bicchieri, C., & Dimant, E. (2023). Norm-nudging: Harnessing social expectations for behavior change. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4418351.
Bicchieri, C., Dimant, E., Gächter, S., & Nosenzo, D. (2022). Social proximity and the erosion of norm compliance. *Games and Economic Behavior, 132*, 59–72. https://doi.org/10.1016/j.geb.2021.11.012.
Bitterly, T. B., & Schweitzer, M. E. (2019). The impression management benefits of humorous self-disclosures: How humor influences perceptions of veracity. *Organizational Behavior and Human Decision Processes, 151*, 73–89.
Biziou-van-Pol, L., Haenen, J., Novaro, A., Occhipinti Liberman, A., & Capraro, V. (2015). Does telling white lies signal pro-social preferences? *Judgment and Decision Making, 10*(6), 538–548. https://doi.org/10.1017/S1930297500006987.
Blanke, O. (2012). Multisensory brain mechanisms of bodily self-consciousness. *Nature Reviews. Neuroscience, 13*, 556–571. https://doi.org/10.1038/nrn3292.
Blanke, O., & Metzinger, T. (2009). Full-body illusions and minimal phenomenal selfhood. *Trends in Cognitive Sciences, 13*, 7–13. https://doi.org/10.1016/j.tics.2008.10.003.
Blazquiz-Pulido, J. F., Polonio, L., & Bilancini, E. (2024). Who's the deceiver? Identifying deceptive intentions in communication. *Games and Economic Behavior, 145*, 451–466. https://doi.org/10.1016/j.geb.2024.02.006.
Boffey, D. (2018). *EU border "lie detector" system criticised as pseudoscience*. The Guardian.
Bolton, G., Dimant, E., & Schmidt, U. (2021). Observability and social image: On the robustness and fragility of reciprocity. *Journal of Economic Behavior & Organization, 191*, 946–964. https://doi.org/10.1016/j.jebo.2021.09.018.
Bourdage, J. S., Roulin, N., & Tarraf, R. (2018). I (might be) just that good: Honest and deceptive impression management in employment interviews. *Personnel Psychology, 71*(4), 597–632.
Braithwaite, J. (2002). *Restorative justice & responsive regulation*. Oxford University Press.
Brañas-Garza, P., Capraro, V., & Rascón-Ramírez, E. (2018). Gender differences in altruism on mechanical turk: Expectations and actual behaviour. *Economics Letters, 170*, 19–23. https://doi.org/10.1016/j.econlet.2018.05.022.
Brimbal, L., Jones, A. M., & Quinby, E. A. (2023). Does telling a story in reverse elicit cues to deceit? A replication and extension of Vrij, Leal, Mann and Fisher (2012). *Legal and Criminological Psychology*, 1–24. https://doi.org/10.1111/lcrp.12252.

Buckenmaier, J., Dimant, E., Posten, A.-C., & Schmidt, U. (2021). Efficient institutions and effective deterrence: On timing and uncertainty of formal sanctions. *Journal of Risk and Uncertainty, 62*(2), 177–201. https://doi.org/10.1007/s11166-021-09352-x.
Bursztyn, L., Egorov, G., & Fiorin, S. (2020). From extreme to mainstream: The erosion of social norms. *American Economic Review, 110*(11), 3522–3548. https://doi.org/10.1257/aer.20171175.
Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin, 125*(3), 367–383. https://doi.org/10.1037/0033-2909.125.3.367.
Cagala, T., Glogowsky, U., Rincke, J., & Schudy, S. (2024). Commitment requests do not affect truth-telling in laboratory and online experiments. *Games and Economic Behavior, 143*, 179–190. https://doi.org/10.1016/j.geb.2023.11.014.
Cai, H., Zhu, J., & Yu, Y. (2020). Robust prediction of individual personality from brain functional connectome. *Social Cognitive and Affective Neuroscience, 15*(3), 359–369. https://doi.org/10.1093/scan/nsaa044.
Caldwell, R. L. (1986). The deceptive use of reputation by stomatopods. In R. W. Mitchell, & N. S. Thompson (Eds.). *Deception: Perspectives on human and nonhuman deceit* (pp. 129–145) State University of New York Press.
Calvano, E., Calzolari, G., Denicolò, V., Harrington, J. E., & Pastorello, S. (2020). Protecting consumers from collusive prices due to AI. *Science, 370*(6520), 1040–1042. https://doi.org/10.1126/science.abe3796.
Cao, H., Plichta, M. M., Schäfer, A., Haddad, L., Grimm, O., Schneider, M., ... Tost, H. (2014). Test–retest reliability of fMRI-based graph theoretical properties during working memory, emotion processing, and resting state. *Neuroimage, 84*, 888–900. https://doi.org/10.1016/j.neuroimage.2013.09.013.
Cappelen, A. W., Sørensen, E.Ø., & Tungodden, B. (2013). When do we lie? *Journal of Economic Behavior & Organization, 93*, 258–265. https://doi.org/10.1016/j.jebo.2013.03.037.
Capraro, V. (2017). Does the truth come naturally? Time pressure increases honesty in one-shot deception games. *Economics Letters, 158*, 54–57. https://doi.org/10.1016/j.econlet.2017.06.015.
Capraro, V. (2018). Gender differences in lying in sender-receiver games: A meta-analysis. *Judgment and Decision Making, 13*(4), 345–355. https://doi.org/10.1017/S1930297500009220.
Capraro, V. (2024). The dual-process approach to human sociality: Meta-analytic evidence for a theory of internalized heuristics for self-preservation. *Journal of Personality and Social Psychology, 126*(5), 719–757. https://doi.org/10.1037/pspa0000375.
Capraro, V., Giardini, F., Vilone, D., & Paolucci, M. (2016). Partner selection supported by opaque reputation promotes cooperative behavior. *Physics and Society, 11*(6), 309–317. https://doi.org/10.1017/S1930297500004800.
Capraro, V., Schulz, J., & Rand, D. G. (2019). Time pressure and honesty in a deception game. *Journal of Behavioral and Experimental Economics, 79*, 93–99. https://doi.org/10.1016/j.socec.2019.01.007.
Capraro, V., & Vanzo, A. (2019). The power of moral words: Loaded language generates framing effects in the extreme dictator game. *Judgment and Decision Making, 14*(3), 309–317. https://doi.org/10.1017/S1930297500004356.
Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin, 140*(4), 980–1008. https://doi.org/10.1037/a0035661.
Chalfin, A., & McCrary, J. (2017). Criminal deterrence: A review of the literature. *Journal of Economic Literature, 55*(1), 5–48. https://doi.org/10.1257/jel.20141147.
Childs, J. (2012). Gender differences in lying. *Economics Letters, 114*(2), 147–149. https://doi.org/10.1016/j.econlet.2011.10.006.

Choo, L., Grimm, V., Horváth, G., & Nitta, K. (2019). Whistleblowing and diffusion of responsibility: An experiment. *European Economic Review, 119*, 287–301. https://doi.org/10.1016/j.euroecorev.2019.07.010.

Chowdhury, R. M. M. I., & Fernando, M. (2014). The relationships of empathy, moral identity and cynicism with consumers' ethical beliefs: The mediating role of moral disengagement. *Journal of Business Ethics, 124*(4), 677–694. https://doi.org/10.1007/s10551-013-1896-7.

Christ S.E., Van Essen D.C., Watson J.M., Brubaker L.E., McDermott K.B. The contributions of prefrontal cortex and executive control to deception: evidence from activation likelihood estimate meta-analyses. Cereb Cortex. 2009 Jul;19(7):1557-66. https://doi.org/10.1093/cercor/bhn189. Epub 2008 Nov 2. PMID: 18980948; PMCID: PMC2693617.

Cialdini, R. B., Demaine, L. J., Sagarin, B. J., Barrett, D. W., Rhoads, K., & Winter, P. L. (2006). Managing social norms for persuasive impact. *Social Influence, 1*(1), 3–15. https://doi.org/10.1080/15534510500181459.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology, 55*(1), 591–621. https://doi.org/10.1146/annurev.psych.55.090902.142015.

Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. *Advances in Experimental Social Psychology*. Academic Press, 201–234. https://doi.org/10.1016/S0065-2601(08)60330-5.

Cohen, T. R., Gunia, B. C., Kim-Jun, S. Y., & Murnighan, J. K. (2009). Do groups lie more than individuals? Honesty & deception as a function of strategic self-interest. *Journal of Experimental Social Psychology, 45*, 1321–1324. https://doi.org/10.1016/j.jesp.2009.08.007.

Cohn, A., Gesche, T., & Maréchal, M. A. (2022). Honesty in the digital age. *Management Science, 68*(2), 827–845. https://doi.org/10.1287/mnsc.2021.3985.

Cohn, A., Maréchal, M. A., Tannenbaum, D., & Zünd, C. L. (2019). Civic honesty around the globe. *Science, 365*(6448), 70–73. https://doi.org/10.1126/science.aau8712.

Conrads, J., Irlenbusch, B., Rilke, R. M., & Walkowitz, G. (2013). Lying and team incentives. *Journal of Economic Psychology, 34*, 1–7. https://doi.org/10.1016/j.joep.2012.10.011.

Cooper, B., Cohen, T. R., Huppert, E., Levine, E. E., & Fleeson, W. (2023). Honest behavior: Truth-seeking, belief-speaking, and fostering understanding of the truth in others. *The Academy of Management Annals, 17*(2), 655–683. https://doi.org/10.5465/annals.2021.0209.

de Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nature Reviews. Neuroscience, 7*(3), 242–249. https://doi.org/10.1038/nrn1872.

Debey, E., De Schryver, M., Logan, G. D., Suchotzki, K., & Verschuere, B. (2015). From junior to senior pinocchio: A cross-sectional lifespan investigation of deception. *Acta Psychologica, 160*, 58–68. https://doi.org/10.1016/j.actpsy.2015.06.007.

Decety, J., & Sommerville, J. A. (2003). Shared representations between self and other: A social cognitive neuroscience view. *Trends in Cognitive Sciences, 7*, 527–533. https://doi.org/10.1016/j.tics.2003.10.004.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology, 18*(1), 105–115. https://doi.org/10.1037/h0030644.

Decker, S., Wright, R., & Logie, R. (1993). Perceptual deterrence among active residential burglars: A research note. *Criminology, 31*(1), 135–147.

Deodhar, P., Jain, S., & Siddiqui, J. A. (2022). Prediction of online academic dishonesty using the voting ensemble machine learning method. 2022 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON), pp. 1–5. https://doi.org/10.1109/SMARTGENCON56628.2022.10084129.

DePaulo, B. M., & Kashy, D. A. (1998). Everyday lies in close and casual relationships. *Journal of personality and social psychology, 74*(1), 63.
DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology, 70*(5), 979–995. https://doi.org/10.1037/0022-3514.70.5.979.
Dijk, C., Koenig, B., Ketelaar, T., & de Jong, P. J. (2011). Saved by the blush: Being trusted despite defecting. *Emotion, 11*(2), 313–319. https://doi.org/10.1037/a0022774.
Dimant, E. (2023). Beyond average: A method for measuring the tightness, looseness, and polarization of social norms. *Economics Letters, 233*, 111417. https://doi.org/10.1016/j.econlet.2023.111417.
Dimant, E. (2024). Hate trumps love: The impact of political polarization on social preferences. *Management Science, 70*(1), 1–31.
Dimant, E., Galeotti, F., & Villeval, M. C. (2024). Motivated information acquisition and social norm formation. *European Economic Review, 167*, 104778. https://doi.org/10.1016/j.euroecorev.2024.104778.
Dimant, E., & Gesche, T. (2023). Nudging enforcers: How norm perceptions and motives for lying shape sanctions. *PNAS Nexus, 2*(7), https://doi.org/10.1093/pnasnexus/pgad224.
Dimant, E., & Shalvi, S. (2022). Meta-nudging honesty: Past, present, and future of the research frontier. *Current Opinion in Psychology, 47*, 101426. https://doi.org/10.1016/j.copsyc.2022.101426.
Dimant, E., van Kleef, G. A., & Shalvi, S. (2020). Requiem for a nudge: Framing effects in nudging honesty. *Journal of Economic Behavior & Organization, 172*, 247–266. https://doi.org/10.1016/j.jebo.2020.02.015.
Dogan, A., Morishima, Y., Heise, F., Tanner, C., Gibson, R., Wagner, A. F., & Tobler, P. N. (2016). Prefrontal connections express individual differences in intrinsic resistance to trading off honesty values against economic benefits. *Scientific Reports, 6*(1), 33263. https://doi.org/10.1038/srep33263.
Doñate-Buendía, A., García-Gallego, A., & Petrović, M. (2022). Gender and other moderators of giving in the dictator game: A meta-analysis. *Journal of Economic Behavior & Organization, 198*, 280–301. https://doi.org/10.1016/j.jebo.2022.03.031.
Doob, A. N., & Webster, C. M. (2003). Sentence severity and crime: Accepting the null hypothesis. *Crime and Justice, 30*, 143–195. https://doi.org/10.1086/652230.
Dores Cruz, T. D., Nieper, A. S., Testori, M., Martinescu, E., & Beersma, B. (2021). An integrative definition and framework to study gossip. *Group & Organization Management, 46*(2), 252–285. https://doi.org/10.1177/1059601121992887.
Dores Cruz, T. D., Thielmann, I., Columbus, S., Molho, C., Wu, J., Righetti, F., ... Balliet, D. (2021). Gossip and reputation in everyday life. *Philosophical Transactions of the Royal Society B: Biological Sciences, 376*(1838), https://doi.org/10.1098/rstb.2020.0301.
Dores Cruz, T. D., van der Lee, R., Bechtoldt, M. N., & Beersma, B. (2024). Nasty and noble notes: Interdependence structures drive self-serving gossip. *Personality and Social Psychology Bulletin, 50*(11), 1596–1612. https://doi.org/10.1177/01461672231171054.
Dorrough, A. R., Köbis, N., Irlenbusch, B., Shalvi, S., & Glöckner, A. (2023). Conditional bribery: Insights from incentivized experiments across 18 nations. *Proceedings of the National Academy of Sciences, 120*(18), https://doi.org/10.1073/pnas.2209731120.
Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science, 3*(5), 562–571. https://doi.org/10.1177/1948550611430272.
Dreber, A., & Johannesson, M. (2008). Gender differences in deception. *Economics Letters, 99*(1), 197–199. https://doi.org/10.1016/j.econlet.2007.06.027.
Dunning, D. (2015). Motivational theories. In B. Gawronski, & G. V. Bodenhausen (Eds.). *Theory and explanation in social psychology* (pp. 108–131) Guilford Press.

Dupont, L., Santangelo, V., Azevedo, R. T., Panasiti, M. S., & Aglioti, S. M. (2023). Reputation risk during dishonest social decision-making modulates anterior insular and cingulate cortex activity and connectivity. *Communications Biology, 6*(1), 475. https://doi.org/10.1038/s42003-023-04827-w.
Duval, S., & Wicklund, R. A. (1973). Effects of objective self-awareness on attribution of causality. *Journal of Experimental Social Psychology, 9*(1), 17–31. https://doi.org/10.1016/0022-1031(73)90059-0.
Effron, D. A., & Helgason, B. A. (2023). *Moral inconsistency. Advances in experimental social psychology*. Academic Press, 1–72. https://doi.org/10.1016/bs.aesp.2022.11.001.
Elaad, E., Kochav, R., & Elkouby, T. (2024). Lying about money and game points by men and women and its relation to the self-reported lying scale. *Frontiers in Psychology, 14*. https://doi.org/10.3389/fpsyg.2023.1304237.
Elliot, A. J., & Devine, P. G. (1994). On the motivational nature of cognitive dissonance: Dissonance as psychological discomfort. *Journal of Personality and Social Psychology, 67*(3), 382–394. https://doi.org/10.1037/0022-3514.67.3.382.
Engel, C. (2011). Dictator games: A meta study. *Experimental Economics, 14*(4), 583–610. https://doi.org/10.1007/s10683-011-9283-7.
Engelmann, J. M., & Herrmann, E. (2016). Chimpanzees trust their friends. *Current Biology, 26*(2), 252–256. https://doi.org/10.1016/j.cub.2015.11.037.
Erat, S., & Gneezy, U. (2012). White lies. *Management Science, 58*(4), 723–733. https://doi.org/10.1287/mnsc.1110.1449.
Erev, I., Hreib, M., & Teodorescu, K. (2024). Big data without big brothers: The potential of gentle rule enforcement. In J. Glückler, & R. Panitz (Vol. Eds.), *Knowledge and digital technology. Knowledge and space. vol. 19. Knowledge and digital technology. Knowledge and space* (pp. 225–237). Cham: Springer. https://doi.org/10.1007/978-3-031-39101-9_12.
European Commission. (2011). *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Youth Opportunities Initiative*. https://www.voced.edu.au/content/ngv:49984.
Ezquerra, L., Kolev, G. I., & Rodriguez-Lara, I. (2018). Gender differences in cheating: Loss vs. gain framing. *Economics Letters, 163*, 46–49. https://doi.org/10.1016/j.econlet.2017.11.016.
Feess, E., Schildberg-Hörisch, H., Schramm, M., & Wohlschlegel, A. (2018). The impact of fine size and uncertainty on punishment and deterrence: Theory and evidence from the laboratory. *Journal of Economic Behavior & Organization, 149*, 58–73. https://doi.org/10.1016/j.jebo.2018.02.021.
Fehr, E., & Schurtenberger, I. (2018). Normative foundations of human cooperation. *Nature human behaviour, 2*(7), 458–468.
Fehr, D., & Sutter, M. (2019). Gossip and the efficiency of interactions. *Games and Economic Behavior, 113*, 448–460. https://doi.org/10.1016/j.geb.2018.10.003.
Feingold, Z. R. (2021). The stigma of incarceration experience: A systematic review. *Psychology, Public Policy, and Law, 27*(4), 550–569. https://doi.org/10.1037/law0000319.
Feldman, Y. (2018). *The law of good people*. Cambridge University Press. https://doi.org/10.1017/9781316480328.
Feruglio, S., Panasiti, M. S., Crescentini, C., Aglioti, S. M., & Ponsi, G. (2024). Training the moral self: An 8-week mindfulness meditation program leads to reduced dishonest behavior and increased regulation of interoceptive awareness. *Mindfulness, 14*(11), 2757–2779. https://doi.org/10.1007/s12671-023-02233-1.
Festinger, L. (1962). Cognitive dissonance. *Scientific American, 207*(4), 93–106. https://doi.org/10.1038/scientificamerican1062-93.
Fiedler, S., & Glöckner, A. (2015). Attention and moral behavior. *Current Opinion in Psychology, 6*, 139–144. https://doi.org/10.1016/j.copsyc.2015.08.008.
Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., ... Constable, R. T. (2015). Functional connectome fingerprinting: Identifying individuals

using patterns of brain connectivity. *Nature Neuroscience, 18*(11), 1664–1671. https://doi.org/10.1038/nn.4135.
Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise: An experimental study on cheating. *Journal of the European Economic Association, 11*(3), 525–547. https://doi.org/10.1111/jeea.12014.
Fleeson, W., & Jayawickreme, E. (2021). *Whole traits: Revealing the social-cognitive mechanisms constituting personality's central variable. Advances in Experimental Social Psychology*Elsevier Academic Press, 69–128. https://doi.org/10.1016/bs.aesp.2020.11.002.
Fosgaard, T., Jacobsen, C., & Street, C. (2021). The heterogeneous processes of cheating: Attention evidence from two eye tracking experiments. *Journal of Behavioral Decision Making, 34*(1), 131–139. https://doi.org/10.1002/bdm.2200.
Frey, B. S., & Jegen, R. (2001). Motivation crowding theory. *Journal of Economic Surveys, 15*(5), 589–611. https://doi.org/10.1111/1467-6419.00150.
Friesen, L., & Gangadharan, L. (2012). Individual level evidence of dishonesty and the gender effect. *Economics Letters, 117*(3), 624–626. https://doi.org/10.1016/j.econlet.2012.08.005.
Frisanco, A., Biella, M., Brambilla, M., & Kret, M. E. (2023). All that meets the eye: The contribution of reward processing and pupil mimicry on pupillary reactions to facial trustworthiness. *Current Psychology, 42*(14), 11685–11692. https://doi.org/10.1007/s12144-021-02486-w.
Fulham, N. M., Krueger, K. L., & Cohen, T. R. (2022). Honest feedback: Barriers to receptivity and discerning the truth in feedback. *Current Opinion in Psychology, 46*, 101405. https://doi.org/10.1016/j.copsyc.2022.101405.
Furr, R. M., & Funder, D. C. (2021). *Persons, situations, and person–situation interactions Handbook of personality: Theory and research* (4th ed.). The Guilford Press, 667–685.
Gächter, S., & Schulz, J. F. (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature, 531*(7595), 496–499. https://doi.org/10.1038/nature17160.
Ganis, G. (2014). Investigating deception and deception detection with brain stimulation methods. In P. A. Granhag, A. Vrij, & B. Verschuere (Eds.). *Detecting deception* (pp. 253–268) (1st ed.). Wiley. https://doi.org/10.1002/9781118510001.ch11.
Gelfand, M., Li, R., Stamkou, E., Pieper, D., Denison, E., Fernandez, J., ... Dimant, E. (2022). Persuading republicans and democrats to comply with mask wearing: An intervention tournament. *Journal of Experimental Social Psychology, 101*, 104299. https://doi.org/10.1016/j.jesp.2022.104299.
Gerlach, P., Teodorescu, K., & Hertwig, R. (2019). The truth about lies: A meta-analysis on dishonest behavior. *Psychological Bulletin, 145*(1), 1–44. https://doi.org/10.1037/bul0000174.
Gerlach, P., & Teodorescu, K. (2022). Measuring dishonest behavior: Hidden dimensions that matter. *Current Opinion in Psychology, 47*, 101408. https://doi.org/10.1016/j.copsyc.2022.101408.
Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review, 95*(1), 384–394. https://doi.org/10.1257/0002828053828662.
Gneezy, U., Imas, A., & Madarász, K. (2014). Conscience accounting: Emotion dynamics and social behavior. *Management Science, 60*(11), 2645–2658. https://doi.org/10.1287/mnsc.2014.1942.
Gneezy, U., Kajackaite, A., & Sobel, J. (2018). Lying aversion and the size of the lie. *American Economic Review, 108*(2), 419–453. https://doi.org/10.1257/aer.20161553.
Gneezy, U., & Rustichini, A. (2000). A fine is a price. *The Journal of Legal Studies, 29*(1), 1–17. https://doi.org/10.1086/468061.
Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research, 35*(3), 472–482. https://doi.org/10.1086/586910.

Graf, C., Pondorfer, A., & Schulz, J. (2024). Culture and gender differences in honesty. (Working Paper).
Greene, J. D., & Paxton, J. M. (2009). Patterns of neural activity associated with honest and dishonest moral decisions. *Proceedings of the National Academy of Sciences, 106*(30), 12506–12511. https://doi.org/10.1073/pnas.0900152106.
Greenwald, A. G. (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist, 35*(7), 603–618. https://doi.org/10.1037/0003-066X.35.7.603.
Grosch, K., & Rau, H. A. (2017). Gender differences in honesty: The role of social value orientation. *Journal Of Economic Psychology, 62*, 258–267. https://doi.org/10.1016/j.joep.2017.07.008.
Gross, J., & de Dreu, C. K. W. (2021). Rule following mitigates collaborative cheating and facilitates the spreading of honesty within groups. *Personality and Social Psychology Bulletin, 47*(3), 395–409. https://doi.org/10.1177/0146167220927195.
Gunia, B. C., Wang, L., Huang, L., Wang, J., & Murnighan, J. K. (2012). Contemplation and conversation: Subtle influences on moral decision making. *Academy of Management Journal, 55*(1), 13–33. https://doi.org/10.5465/amj.2009.0873.
Haggard, P. (2017). Sense of agency in the human brain. *Nature Reviews. Neuroscience, 18*(4), 196–207. https://doi.org/10.1038/nrn.2017.14.
Halevy, R., Shalvi, S., & Verschuere, B. (2014). Being honest about dishonesty: Correlating self-reports and actual lying. *Human Communication Research, 40*(1), 54–72. https://doi.org/10.1111/hcre.12019.
Hallsworth, M., List, J. A., Metcalfe, R. D., & Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics, 148*, 14–31. https://doi.org/10.1016/j.jpubeco.2017.02.003.
Hardin, R. (2002). *Trust and trustworthiness*. Russell Sage Foundation.
Hare, B., Wobber, V., & Wrangham, R. (2012). The self-domestication hypothesis: Evolution of bonobo psychology is due to selection against aggression. *Animal Behaviour, 83*(3), 573–585. https://doi.org/10.1016/j.anbehav.2011.12.007.
Hatta, H., Ueda, R., Ashida, H., & Abe, N. (2022). Are implicit attitudes toward dishonesty associated with self-serving dishonesty? Implications for the reliability of the IAT. *Journal of Experimental Social Psychology, 100*, 104285. https://doi.org/10.1016/j.jesp.2022.104285.
Heck, D. W., Thielmann, I., Moshagen, M., & Hilbig, B. E. (2018). Who lies? A large-scale reanalysis linking basic personality traits to unethical decision making. *Judgment and Decision Making, 13*(4), 356–371. https://doi.org/10.1017/S1930297500009232.
Heinicke, F., Rosenkranz, S., & Weitzel, U. (2019). The effect of pledges on the distribution of lying behavior: An online experiment. *Journal Of Economic Psychology, 73*, 136–151. https://doi.org/10.1016/j.joep.2019.05.006.
Helgason, B. A., & Effron, D. A. (2022). From critical to hypocritical: Counterfactual thinking increases partisan disagreement about media hypocrisy. *Journal of Experimental Social Psychology, 101*, 104308.
Hertwig, R., & Mazar, N. (2022). Toward a taxonomy and review of honesty interventions. *Current Opinion in Psychology, 47*, 101410. https://doi.org/10.1016/j.copsyc.2022.101410.
Hildreth, J. A. D., & Anderson, C. (2018). Does loyalty trump honesty? Moral judgments of loyalty-driven deceit. *Journal of Experimental Social Psychology, 79*, 87–94.
Hilbig, B. E. (2022). Personality and behavioral dishonesty. *Current Opinion in Psychology, 47*, 101378. https://doi.org/10.1016/j.copsyc.2022.101378.
Hilbig, B. E., Moshagen, M., Thielmann, I., & Zettler, I. (2022). Making rights from wrongs: The crucial role of beliefs and justifications for the expression of aversive personality. *Journal of Experimental Psychology: General, 151*(11), 2730–2755. https://doi.org/10.1037/xge0001232.

Hilbig, B. E., Thielmann, I., & Heck, D. W. (2024). Filling in the missing pieces: Personality traits (un)related to dishonest behavior. *European Journal of Personality. Advance online publication*. https://doi.org/10.1177/08902070241293621.

Hilbig, B. E., & Thielmann, I. (2017). Does everyone have a price? On the role of payoff magnitude for ethical decision making. *Cognition, 163*, 15–25.

Hilbig, B. E., Thielmann, I., Zettler, I., & Moshagen, M. (2023). The dispositional essence of proactive social preferences: The dark core of personality vis-à-vis 58 traits. *Psychological Science, 34*(2), 201–220. https://doi.org/10.1177/09567976221116893.

Hilbig, B. E., & Zettler, I. (2015). When the cat's away, some mice will play: A basic trait account of dishonest behavior. *Journal of Research in Personality, 57*, 72–88. https://doi.org/10.1016/j.jrp.2015.04.003.

Hochman, G., & Erev, I. (2013). The partial-reinforcement extinction effect and the contingent-sampling hypothesis. *Psychonomic Bulletin & Review, 20*(6), 1336–1342. https://doi.org/10.3758/s13423-013-0432-1.

Hochman, G., Glöckner, A., Fiedler, S., & Ayal, S. (2016). I can see it in your eyes": Biased processing and increased arousal in dishonest responses. *Journal of Behavioral Decision Making, 29*(2–3), 322–335. https://doi.org/10.1002/bdm.1932.

Hochman, G., Peleg, D., Ariely, D., & Ayal, S. (2021). Robin hood meets pinocchio: Justifications increase cheating behavior but decrease physiological tension. *Journal of Behavioral and Experimental Economics, 92*, 101699. https://doi.org/10.1016/j.socec.2021.101699.

Hofmann, W., & Grigoryan, L. (2023). *The social psychology of everyday life. Advances in experimental social psychology Vol. 68. Advances in experimental social psychology* Academic Press Inc.77–137. https://doi.org/10.1016/bs.aesp.2023.06.001.

Hohenstein, J., & Jung, M. (2020). AI as a moral crumple zone: The effects of AI-mediated communication on attribution and trust. *Computers in Human Behavior, 106*, 106190. https://doi.org/10.1016/j.chb.2019.106190.

Huber, C., Litsios, C., Nieper, A., & Promann, T. (2023). On social norms and observability in (dis)honest behavior. *Journal of Economic Behavior & Organization, 212*, 1086–1099. https://doi.org/10.1016/j.jebo.2023.06.019.

Jacquemet, N., Luchini, S., Rosaz, J., & Shogren, J. F. (2019). Truth telling under oath. *Management Science, 65*(1), 426–438. https://doi.org/10.1287/mnsc.2017.2892.

Jang, H., Boesch, C., Mundry, R., Ban, S. D., & Janmaat, K. R. L. (2019). Travel linearity and speed of human foragers and chimpanzees during their daily search for food in tropical rainforests. *Scientific Reports, 9*(1), 11066. https://doi.org/10.1038/s41598-019-47247-9.

Janis, I. L. (2008). Groupthink. *IEEE Engineering Management Review, 36*(1), 36.

Jayawickreme, E., Fleeson, W., Miller, C. B., Batson, C. D., Sreenivasan, G., Levy, N., ... Cameron, C. D. (2017). *Does whole trait theory work for the virtues? Moral psychology, Volume 5: Virtue and character: Vol. V*. MIT Press, 75–103.

Johnson, A. K., Barnacz, A., Yokkaichi, T., Rubio, J., Racioppi, C., Shackelford, T. K., ... Keenan, J. P. (2005). Me, myself, and lie: The role of self-awareness in deception. *Personality and Individual Differences, 38*(8), 1847–1853. https://doi.org/10.1016/j.paid.2004.11.013.

Johnson, D. J., Cheung, F., & Donnellan, M. B. (2014). Does cleanliness influence moral judgments? A direct replication of Schnall, Benton, and Harvey (2008). *Social Psychology, 45*(3), 209–215. https://doi.org/10.1027/1864-9335/a000186.

Ju, U. (2023). Task and resting-state functional connectivity predict driving violations. *Brain Sciences, 13*(9), 1236. https://doi.org/10.3390/brainsci13091236.

Kajackaite, A., & Gneezy, U. (2017). Incentives and cheating. *Games and Economic Behavior, 102*, 433–444. https://doi.org/10.1016/j.geb.2017.01.015.

Kashy, D. A., & DePaulo, B. M. (1996). Who lies? *Journal of Personality and Social Psychology, 70*(5), 1037–1051. https://doi.org/10.1037/0022-3514.70.5.1037.
Kastlunger, B., Kirchler, E., Mittone, L., & Pitters, J. (2009). Sequences of audits, tax compliance, and taxpaying strategies. *Journal Of Economic Psychology, 30*(3), 405–418. https://doi.org/10.1016/j.joep.2008.10.004.
Keep, W. (2009). Furthering organizational priorities with less than truthful behavior: A call for additional tools. *Journal of Business Ethics, 86*, 81–90.
Kelley, H. H., Holmes, J. G., Kerr, N. L., Reis, H. T., Rusbult, C. E., & van Lange, P. A. M. (2003). *An Atlas of Interpersonal Situations*. Cambridge University Press. https://doi.org/10.1017/CBO9780511499845.
Kennedy, J. A., & Kray, L. J. (2022). Gender similarities and differences in dishonesty. *Current Opinion in Psychology, 48*, 101461. https://doi.org/10.1016/j.copsyc.2022.101461.
Kennedy, J. A., & Schweitzer, M. E. (2018). Building trust by tearing others down: When accusing others of unethical behavior engenders trust. *Organizational Behavior and Human Decision Processes, 149*, 111–128. https://doi.org/10.1016/j.obhdp.2018.10.001.
Kettle, S., Hernandez, M., Sanders, M., Hauser, O., & Ruda, S. (2017). Failure to CAPTCHA attention: Null results from an honesty priming experiment in guatemala. *Behavioral Sciences, 7*(2), 28. https://doi.org/10.3390/bs7020028.
Klein, S. A., Thielmann, I., Hilbig, B. E., & Zettler, I. (2017). Between me and we: The importance of self-profit versus social justifiability for ethical decision making. *Judgment and Decision Making, 12*(6), 563–571. https://doi.org/10.1017/S1930297500006690.
Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science (New York, N. Y.), 314*(5800), 829–832. https://doi.org/10.1126/science.1129156.
Köbis, N., Bonnefon, J.-F., & Rahwan, I. (2021). Bad machines corrupt good morals. *Nature Human Behaviour, 5*(6), 679–685. https://doi.org/10.1038/s41562-021-01128-2.
Köbis, N., Rahwan, Z., Bersch, C., Ajaj, T., Bonnefon, J.F., & Rahwan, I. (2024). Experimental evidence that delegating to intelligent machines can increase dishonest behaviour (No. dnjgz). Center for Open Science. (originally cited as Rahwan et al., 2024).
Köbis, N., Starke, C., & Rahwan, I. (2022). The promise and perils of using artificial intelligence to fight corruption. *Nature Machine Intelligence, 4*(5), 418–424. https://doi.org/10.1038/s42256-022-00489-1.
Köbis, N. C., Verschuere, B., Bereby-Meyer, Y., Rand, D., & Shalvi, S. (2019). Intuitive honesty versus dishonesty: Meta-analytic evidence. *Perspectives on Psychological Science, 14*(5), 778–796. https://doi.org/10.1177/1745691619851778.
Kocher, M. G., Schudy, S., & Spantig, L. (2018). I Lie? We Lie! Why? Experimental evidence on a dishonesty shift in groups. *Management Science, 64*(9), 3995–4008. https://doi.org/10.1287/mnsc.2017.2800.
Kohn, M. L., & Schooler, C. (1983). *Work and personality: An inquiry into the impact of social stratification*. Ablex Publishing Corporation.
Koster, R., Balaguer, J., Tacchetti, A., Weinstein, A., Zhu, T., Hauser, O., ... Summerfield, C. (2022). Human-centred mechanism design with democratic AI. *Nature Human Behaviour, 6*(10), 1398–1407. https://doi.org/10.1038/s41562-022-01383-x.
Kret, M. E., & Akyüz, R. (2022). Mimicry eases prediction and thereby smoothens social interactions. *Cognition and Emotion, 36*(5), 794–798. https://doi.org/10.1080/02699931.2022.2110452.
Kret, M. E., & de Dreu, C. K. W. (2019). The power of pupil size in establishing trust and reciprocity. *Journal of Experimental Psychology: General, 148*(8), 1299–1311. https://doi.org/10.1037/xge0000508.
Kret, M. E., Fischer, A. H., & de Dreu, C. K. W. (2015). Pupil mimicry correlates with trust in in-group partners with dilating pupils. *Psychological Science, 26*(9), 1401–1410. https://doi.org/10.1177/0956797615588306.

Kreysa, H., Kessler, L., & Schweinberger, S. R. (2016). Direct speaker gaze promotes trust in truth-ambiguous statements. *PLoS ONE, 11*(9), e0162291. https://doi.org/10.1371/journal.pone.0162291.

Kristal, A. S., Whillans, A. V., Bazerman, M. H., Gino, F., Shu, L. L., Mazar, N., & Ariely, D. (2020). Signing at the beginning versus at the end does not decrease dishonesty. *Proceedings of the National Academy of Sciences, 117*(13), 7103–7107. https://doi.org/10.1073/pnas.1911695117.

Krupka, E. L., & Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association, 11*(3), 495–524. https://doi.org/10.1111/jeea.12006.

Kubbe, I., & Merkle, O. (Eds.). (2022). *Norms, gender and corruption: Understanding the Nexus*. Edward Elgar Publishing.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480.

Laustsen, L., & Petersen, M. B. (2018). When the party decides: The effects of facial competence and dominance on internal nominations of political candidates. *Evolutionary Psychology, 16*(2), https://doi.org/10.1177/1474704917732005.

le Maux, B., & Necker, S. (2023). Honesty nudges: Effect varies with content but not with timing. *Journal of Economic Behavior & Organization, 207*, 433–456. https://doi.org/10.1016/j.jebo.2023.01.011.

Lee, S. W. S., & Schwarz, N. (2010). Dirty hands and dirty mouths: Embodiment of the moral-purity metaphor is specific to the motor modality involved in moral transgression. *Psychological Science, 21*(10), 1423–1425. https://doi.org/10.1177/0956797610382788.

Leib, M., Köbis, N., Rilke, R. M., Hagens, M., & Irlenbusch, B. (2024). Corrupted by algorithms? How AI-generated and human-written advice shape (dis) honesty. *The Economic Journal, 134*(658), 766–784.

Leib, M., Köbis, N., Soraperra, I., Weisel, O., & Shalvi, S. (2021). Collaborative dishonesty: A meta-analytic review. *Psychological Bulletin, 147*(12), 1241–1268. https://doi.org/10.1037/bul0000349.

Leib, M., Pittarello, A., Gordon-Hecker, T., Shalvi, S., & Roskes, M. (2019). Loss framing increases self-serving mistakes (but does not alter attention). *Journal of Experimental Social Psychology, 85*, 103880. https://doi.org/10.1016/j.jesp.2019.103880.

Lenggenhager, B., Azevedo, R. T., Mancini, A., & Aglioti, S. M. (2013). Listening to your heart and feeling yourself: Effects of exposure to interoceptive signals during the ultimatum game. *Experimental Brain Researcg, 230*, 233–241. https://doi.org/10.1007/s00221-013-3647-5.

Levenson, R. W., & Gottman, J. M. (1983). Marital interaction: Physiological linkage and affective exchange. *Journal of Personality and Social Psychology, 45*(3), 587–597. https://doi.org/10.1037/0022-3514.45.3.587.

Levine, T. R. (2014). Truth-default theory (TDT). *Journal of Language and Social Psychology, 33*(4), 378–392. https://doi.org/10.1177/0261927X14535916.

Levine, E. E., & Cohen, T. R. (2018). You can handle the truth: Mispredicting the consequences of honest communication. *Journal of Experimental Psychology: General, 147*(9), 1400.

Levine, E., Hart, J., Moore, K., Rubin, E., Yadav, K., & Halpern, S. (2018). The surprising costs of silence: Asymmetric preferences for prosocial lies of commission and omission. *Journal of personality and social psychology, 114*(1), 29.

Levine, E. E., Roberts, A. R., & Cohen, T. R. (2020). Difficult conversations: Navigating the tension between honesty and benevolence. *Current Opinion in Psychology, 31*, 38–43.

Levine, E. E., & Duncan, S. (2022). Deception and the marketplace of ideas. *Consumer Psychology Review, 5*(1), 33–50.

Levine, E. E., & Lupoli, M. J. (2022). Prosocial lies: Causes and consequences. *Current Opinion in Psychology, 43*, 335–340. https://doi.org/10.1016/j.copsyc.2021.08.006.

Levine, E. E., & Schweitzer, M. E. (2014). Are liars ethical? On the tension between benevolence and honesty. *Journal of Experimental Social Psychology, 53*, 107–117.

Levine, E. E., & Schweitzer, M. E. (2015). Prosocial lies: When deception breeds trust. *Organizational Behavior and Human Decision Processes, 126*, 88–106.

Levine, E. E. (2022). Community standards of deception: Deception is perceived to be ethical when it prevents unnecessary harm. *Journal of Experimental Psychology: General, 151*(2), 410–436. https://doi.org/10.1037/xge0001081.

Li, Y., Chen, Z., Liu, X., & Qi, Y. (2022). Perceiving the facial trustworthiness: Facial age, emotional expression, and attractiveness. *Quarterly Journal of Experimental Psychology, 75*(5), 818–829. https://doi.org/10.1177/17470218211047176.

Li, N., Ma, N., Liu, Y., He, X.-S., Sun, D.-L., Fu, X.-M., ... Zhang, D.-R. (2013). Resting-state functional connectivity predicts impulsivity in economic decision-making. *The Journal of Neuroscience, 33*(11), 4886–4895. https://doi.org/10.1523/JNEUROSCI.1342-12.2013.

Lilleholt, L., Schild, C., & Zettler, I. (2020). Not all computerized cheating tasks are equal: A comparison of computerized and non-computerized versions of a cheating task. *Journal Of Economic Psychology, 78*, 102270. https://doi.org/10.1016/j.joep.2020.102270.

Lima, G., Grgić-Hlača, N., & Cha, M. (2021). Human perceptions on moral responsibility of AI: A case study in AI-assisted bail decision-making. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–17. https://doi.org/10.1145/3411764.3445260.

Liuzza, M. T., Candidi, M., & Aglioti, S. M. (2011). Do not resonate with actions: Sentence polarity modulates cortico-spinal excitability during action-related sentence reading. *PLoS ONE, 6*(2), e16855. https://doi.org/10.1371/journal.pone.0016855.

Liuzza, M. T., Candidi, M., Sforza, A. L., & Aglioti, S. M. (2015). Harm avoiders suppress motor resonance to observed immoral actions. *Social Cognitive and Affective Neuroscience, 10*(1), 72–77. https://doi.org/10.1093/scan/nsu025.

Lohse, T., & Qari, S. (2021). Gender differences in face-to-face deceptive behavior. *Journal of Economic Behavior & Organization, 187*, 1–15. https://doi.org/10.1016/j.jebo.2021.03.026.

Lohse, T., Simon, S. A., & Konrad, K. A. (2018). Deception under time pressure: Conscious decision or a problem of awareness? *Journal of Economic Behavior & Organization, 146*, 31–42. https://doi.org/10.1016/j.jebo.2017.11.026.

Lois, G., & Wessa, M. (2021). Honest mistake or perhaps not: The role of descriptive and injunctive norms on the magnitude of dishonesty. *Journal of Behavioral Decision Making, 34*(1), 20–34. https://doi.org/10.1002/bdm.2196.

López-Pérez, R. (2012). The power of words: A model of honesty and fairness. *Journal Of Economic Psychology, 33*(3), 642–658. https://doi.org/10.1016/j.joep.2011.12.004.

Lundquist, T., Ellingsen, T., Gribbe, E., & Johannesson, M. (2009). The aversion to lying. *Journal of Economic Behavior & Organization, 70*(1–2), 81–92. https://doi.org/10.1016/j.jebo.2009.02.010.

Lupoli, M. J., Levine, E. E., & Greenberg, A. E. (2018). Paternalistic lies. *Organizational Behavior and Human Decision Processes, 146*, 31–50.

Mac Giolla, E., & Luke, T. J. (2021). Does the cognitive approach to lie detection improve the accuracy of human observers? *Applied Cognitive Psychology, 35*(2), 385–392. https://doi.org/10.1002/acp.3777.

Maggian, V., & Villeval, M. C. (2016). Social preferences and lying aversion in children. *Experimental Economics, 19*, 663–685.

Mann, H., Garcia-Rada, X., Hornuf, L., Tafurt, J., & Ariely, D. (2016). Cut from the same cloth: Similarly dishonest individuals across countries. *Journal of Cross-Cultural Psychology, 47*(6), 858–874. https://doi.org/10.1177/0022022116648211.

Martinescu, E., Jansen, W., & Beersma, B. (2021). Negative gossip decreases targets' organizational citizenship behavior by decreasing social inclusion. A multi-method

approach. *Group & Organization Management, 46*(3), 463–497. https://doi.org/10.1177/1059601120986876.
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. https://doi.org/10.1509/jmkr.45.6.633.
McKimmie, B. M. (2015). Cognitive dissonance in groups. *Social and Personality Psychology Compass, 9*(4), 202–212. https://doi.org/10.1111/spc3.12167.
Mead, N. L., Baumeister, R. F., Gino, F., Schweitzer, M. E., & Ariely, D. (2009). Too tired to tell the truth: Self-control resource depletion and dishonesty. *Journal of Experimental Social Psychology, 45*(3), 594–597. https://doi.org/10.1016/j.jesp.2009.02.004.
Meindl, P., Jayawickreme, E., Furr, R. M., & Fleeson, W. (2015). A foundation beam for studying morality from a personological point of view: Are individual differences in moral behaviors and thoughts consistent? *Journal of Research in Personality, 59*, 81–92. https://doi.org/10.1016/j.jrp.2015.09.005.
Merriam-Webster. (n.d.). *Merriam-Webster.com dictionary*. Retrieved from https://www.merriam-webster.com/dictionary/honest. (Accessed 3 December 2024).
Mittone, L. (2006). Dynamic behaviour in tax evasion: An experimental approach. *The Journal of Socio-Economics, 35*(5), 813–835. https://doi.org/10.1016/j.socec.2005.11.065.
Moore, C. (2015). Moral disengagement. *Current Opinion in Psychology, 6*, 199–204. https://doi.org/10.1016/j.copsyc.2015.07.018.
Moro, V., Urgesi, C., Pernigo, S., Lanteri, P., Pazzaglia, M., & Aglioti, S. M. (2008). The neural basis of body form and body action agnosia. *Neuron, 60*(2), 235–246. https://doi.org/10.1016/j.neuron.2008.09.022.
Moshagen, M., Hilbig, B. E., & Zettler, I. (2018). The dark core of personality. *Psychological Review, 125*(5), 656–688. https://doi.org/10.1037/rev0000111.
Moshagen, M., Hilbig, B. E., & Zettler, I. (2024). How and why aversive personality is expressed in political preferences. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspp0000498.
Moshagen, M., Zettler, I., Horsten, L. K., & Hilbig, B. E. (2020). Agreeableness and the common core of dark traits are functionally different constructs. *Journal of Research in Personality, 87*, 103986. https://doi.org/10.1016/j.jrp.2020.103986.
Mouginot, M., Wilson, M. L., Desai, N., & Surbeck, M. (2024). Differences in expression of male aggression between wild bonobos and chimpanzees. *Current Biology, 34*(8), 1780–1785.e4. https://doi.org/10.1016/j.cub.2024.02.071.
Muehlheusser, G., Roider, A., & Wallmeier, N. (2015). Gender differences in honesty: Groups versus individuals. *Economics Letters, 128*, 25–29. https://doi.org/10.1016/j.econlet.2014.12.019.
Muñoz García, A., Gil-Gómez de Liaño, B., & Pascual-Ezama, D. (2021). Gender differences in individual dishonesty profiles. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.728115.
Muris, P., Merckelbach, H., Otgaar, H., & Meijer, E. (2017). The malevolent side of human nature. *Perspectives on Psychological Science, 12*(2), 183–204. https://doi.org/10.1177/1745691616666070.
Nagin, D. S. (2013). Deterrence in the twenty-first Century. *Crime and Justice, 42*(1), 199–263. https://doi.org/10.1086/670398.
Nagin, D. S., Cullen, F. T., & Jonson, C. L. (2009). Imprisonment and reoffending. *Crime and justice, 38*(1), 115–200.
Ng, N. L., Luke, D. M., & Gawronski, B. (2022). Prediction of dishonest behavior using the CNI model of moral dilemma judgment: A registered report. *Comprehensive Results in Social Psychology, 6*(1–3), 85–112. https://doi.org/10.1080/23743603.2022.2102473.

Nieper, A. S., Beersma, B., Dijkstra, M. T. M., & van Kleef, G. A. (2022). When and why does gossip increase prosocial behavior? *Current Opinion in Psychology, 44*, 315–320. https://doi.org/10.1016/j.copsyc.2021.10.009.

Nieper, A. S., Beersma, B., Dijkstra, M. T. M., & Van Kleef, G. A. (2025). When and why does observability increase honesty? The role of gossip and reputational concern. *Judgment and Decision Making, 20*, e5.

Nostro, A. D., Müller, V. I., Varikuti, D. P., Pläschke, R. N., Hoffstaedter, F., Langner, R., … Eickhoff, S. B. (2018). Predicting personality from network-based resting-state functional connectivity. *Brain Structure and Function, 223*(6), 2699–2719. https://doi.org/10.1007/s00429-018-1651-z.

Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science (New York, N. Y.), 314*(5805), 1560–1563. https://doi.org/10.1126/science.1133755.

Panasiti, M. S., Cardone, D., Pavone, E. F., Mancini, A., Merla, A., & Aglioti, S. M. (2016). Thermal signatures of voluntary deception in ecological conditions. *Scientific Reports, 6*, 35174. https://doi.org/10.1038/srep35174.

Panasiti, M. S., Pavone, E. F., Mancini, A., Merla, A., Grisoni, L., & Aglioti, S. M. (2014). The motor cost of telling lies: Electrocortical signatures and personality foundations of spontaneous deception. *Social Neuroscience, 9*(6), 573–589. https://doi.org/10.1080/17470919.2014.934394.

Panasiti, M. S., Pavone, E. F., Merla, A., & Aglioti, S. M. (2011). Situational and dispositional determinants of intentional deceiving. *PLoS ONE, 6*(4), e19465. https://doi.org/10.1371/journal.pone.0019465.

Pang, L., Li, H., Liu, Q., Luo, Y.-J., Mobbs, D., & Wu, H. (2021). Resting-state functional connectivity of social brain regions predicts motivated dishonesty. *NeuroImage*. https://doi.org/10.1101/2021.11.10.468161.

Panizza, F., Dimant, E., Kimbrough, E. O., & Vostroknutov, A. (2024). Measuring norm pluralism and perceived polarization in US politics. *PNAS Nexus, 3*(10), https://doi.org/10.1093/pnasnexus/pgae413.

Park, Y., & Klein, N. (2024). Ghosting: Social rejection without explanation, but not without care. *Journal of Experimental Psychology: General*.

Peer, E., Acquisti, A., & Shalvi, S. (2014). I cheated, but only a little: Partial confessions to unethical behavior. *Journal of Personality and Social Psychology, 106*(2), 202–217. https://doi.org/10.1037/a0035392.

Peer, E., & Feldman, Y. (2021). Honesty pledges for the behaviorally-based regulation of dishonesty. *Journal of European Public Policy, 28*(5), 761–781. https://doi.org/10.1080/13501763.2021.1912149.

Peer, E., Mazar, N., Feldman, Y., & Ariely, D. (2024). How pledges reduce dishonesty: The role of involvement and identification. *Journal of Experimental Social Psychology, 113*, 104614. https://doi.org/10.1016/j.jesp.2024.104614.

Peleg, D., Ayal, S., Ariely, D., & Hochman, G. (2019). The Lie deflator – The effect of polygraph test feedback on subsequent (dis)honesty. *Judgment and Decision Making, 14*(6), 728–738. https://doi.org/10.1017/S1930297500005441.

Peleg, D., Hochman, G., Levine, T., Klar, Y., & Ayal, S. (2021). Revenge is not blind: Testing the ability of retribution to justify dishonesty. *Judgment and Decision Making, 16*(6), 1525–1548. https://doi.org/10.1017/S1930297500008536.

Pennycook, G., Bear, A., Collins, E. T., & Rand, D. G. (2020). The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings. *Management Science, 66*(11), 4944–4957.

Peters, K., & Fonseca, M. A. (2020). Truth, lies, and gossip. *Psychological Science, 31*(6), 702–714. https://doi.org/10.1177/0956797620916708.

Pinker, S. (2012). *The better angels of our nature: Why violence has declined*. Penguin Books.

Pittarello, A., Conte, B., Caserotti, M., Scrimin, S., & Rubaltelli, E. (2017). Emotional intelligence buffers the effect of physiological arousal on dishonesty. *Psychonomic Bulleting & Review, 25*(1), 1–7. https://doi.org/10.3758/s13423-017-1285-9.

Pittarello, A., Frătescu, M., & Mathôt, S. (2019). Visual saliency influences ethical blind spots and (dis)honesty. *Psychonomic Bulletin & Review, 26*(5), 1719–1728. https://doi.org/10.3758/s13423-019-01638-1.

Pittarello, A., Leib, M., Gordon-Hecker, T., & Shalvi, S. (2015). Justifications shape ethical blind spots. *Psychological Science, 26*(6), 794–804. https://doi.org/10.1177/0956797615571018.

Pittarello, A., Motro, D., Rubaltelli, E., & Pluchino, P. (2016). The relationship between attention allocation and cheating. *Psychonomic Bulletin & Review, 23*(2), 609–616. https://doi.org/10.3758/s13423-015-0935-z.

Plonsky, O., Teodorescu, K., & Erev, I. (2015). Reliance on small samples, the wavy recency effect, and similarity-based learning. *Psychological Review, 122*(4), 621–647. https://doi.org/10.1037/a0039413.

Pulfrey, C., & Butera, F. (2013). Why neoliberal values of self-enhancement lead to cheating in higher education. *Psychological Science, 24*(11), 2153–2162. https://doi.org/10.1177/0956797613487221.

Pulfrey, C., & Butera, F. (2016). When and why people don't accept cheating: Self-transcendence values, social responsibility, mastery goals and attitudes towards cheating. *Motivation and Emotion, 40*(3), 438–454. https://doi.org/10.1007/s11031-015-9530-x.

Pulfrey, C., Durussel, K., & Butera, F. (2018). The good cheat: Benevolence and the justification of collective cheating. *Journal of Educational Psychology, 110*(6), 764–784. https://doi.org/10.1037/edu0000247.

Rand, D. G., Brescoll, V. L., Everett, J. A. C., Capraro, V., & Barcelo, H. (2016). Social heuristics and social roles: Intuition favors altruism for women but not for men. *Journal of Experimental Psychology: General, 145*(4), 389–396. https://doi.org/10.1037/xge0000154.

Rauthmann, J. F. (2021). *The handbook of personality dynamics and processes*. Academic Press.

Reinhardt, N., & Reinhard, M. A. (2023). Honesty–humility negatively correlates with dishonesty in romantic relationships. *Journal of Personality and Social Psychology, 125*(4), 925.

Ren, Z., Hart, E., Levine, E. E., & Schweitzer, M. E. (2022). The shared responsibility model of deception. *Current Opinion in Psychology, 48*, 101470. https://doi.org/10.1016/j.copsyc.2022.101470.

Rosenberg, M., Schooler, C., Schoenbach, C., & Rosenberg, F. (1995). Global self-esteem and specific self-esteem: Different concepts, different outcomes. *American Sociological Review, 60*(1), 141. https://doi.org/10.2307/2096350.

Ruffle, B. J., & Tobol, Y. (2014). Honest on Mondays: Honesty and the temporal separation between decisions and payoffs. *European Economic Review, 65*, 126–135. https://doi.org/10.1016/j.euroecorev.2013.11.004.

Ruffle, B. J., & Tobol, Y. (2017). Clever enough to tell the truth. *Experimental Economics, 20*(1), 130–155. https://doi.org/10.1007/s10683-016-9479-y.

Rullo, M., Presaghi, F., Baldner, C., Livi, S., & Butera, F. (2024). *Omertà* in intragroup cheating: The role of ingroup identity in dishonesty and whistleblowing. *Group Processes & Intergroup Relations, 27*(1), 41–61. https://doi.org/10.1177/13684302231164722.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54–67. https://doi.org/10.1006/ceps.1999.1020.

Sachdeva, S., Iliev, R., & Medin, D. L. (2009). Sinning saints and saintly sinners. *Psychological Science, 20*(4), 523–528. https://doi.org/10.1111/j.1467-9280.2009.02326.x.

Salomon, R. (2017). The assembly of the self from sensory and motor foundations. *Social Cognition, 35*(2), 87–106. https://doi.org/10.1521/soco.2017.35.2.87.

Sanitioso, R., Kunda, Z., & Fong, G. T. (1990). Motivated recruitment of autobiographical memories. *Journal of Personality and Social Psychology, 59*(2), 229–241. https://doi.org/10.1037/0022-3514.59.2.229.

Sarzyńska, J., Falkiewicz, M., Riegel, M., Babula, J., Margulies, D. S., Nęcka, E., ... Szatkowska, I. (2017). More intelligent extraverts are more likely to deceive. *PLoS One, 12*(4), e0176591. https://doi.org/10.1371/journal.pone.0176591.

Scattolin, M., Panasiti, M. S., & Aglioti, S. M. (2022a). Morality in the flesh: On the link between bodily self-consciousness, moral identity and (dis) honest behaviour. *Royal Society Open Science, 9*(8), 220061. https://doi.org/10.1098/rsos.220061.

Scattolin, M., Panasiti, M. S., Ho, J. T., Lenggenhager, B., & Aglioti, S. M. (2023). Ownership of the affected leg is further reduced following deceptive behaviors in body integrity dysphoria. *iScience, 26*(9), 107551. https://doi.org/10.1016/j.isci.2023.107551.

Scattolin, M., Panasiti, M. S., Villa, R., & Aglioti, S. M. (2022b). Reduced ownership over a virtual body modulates dishonesty. *iScience, 25*(5), 104320. https://doi.org/10.1016/j.isci.2022.104320.

Schein, E. H. (2010). *Organizational culture and leadershipVol. 4th.* Jossey-Bass.

Schild, C., Moshagen, M., Ścigała, K. A., & Zettler, I. (2020). May the odds — or your personality — be in your favor: Probability of observing a favorable outcome, Honesty-Humility, and dishonest behavior. *Judgment and Decision Making, 15*(4), 600–610. https://doi.org/10.1017/S193029750000752X.

Schnall, S., Benton, J., & Harvey, S. (2008). With a clean conscience: Cleanliness reduces the severity of moral judgments. *Psychological Science, 19*(12), 1219–1222. https://doi.org/10.1111/j.1467-9280.2008.02227.x.

Scholz, D. D., Zimmermann, J., Moshagen, M., Zettler, I., & Hilbig, B. E. (2024). Theoretical and empirical integration of "Dark" traits and socially aversive personality psychopathology. *Journal of Personality Disorders, 38*(3), 241–267. https://doi.org/10.1521/pedi.2024.38.3.241.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science, 18*(5), 429–434. https://doi.org/10.1111/j.1467-9280.2007.01917.x.

Schurr, A., Ritov, I., Kareev, Y., & Avrahami, J. (2012). Is that the answer you had in mind? The effect of perspective on unethical behavior. *Judgment and Decision Making, 7*(6), 679–688.

Schwardmann, P., & Van der Weele, J. (2019). Deception and self-deception. *Nature human behaviour, 3*(10), 1055–1061.

Ścigała, K. A., Schild, C., & Zettler, I. (2020). Dishonesty as a signal of trustworthiness: Honesty-humility and trustworthy dishonesty. *Royal Society Open Science,* 7(10), 200685. https://doi.org/10.1098/rsos.200685.

Ścigała, K. A., Schild, C., Heck, D. W., & Zettler, I. (2019). Who deals with the devil? Interdependence, personality, and corrupted collaboration. *Social Psychological and Personality Science, 10*(8), 1019–1027. https://doi.org/10.1177/1948550618813419.

Seidl, A., Hilbig, B. E., & Thielmann, I. (2024). A partner in crime: Revisiting the dispositional basis of unethical loyalty. *Unpublished Manuscript*.

Seligman, A. B. (2000). *The problem of trust.* Princeton University Press.

Serota, K. B., Levine, T. R., & Boster, F. J. (2010). The prevalence of lying in America: Three studies of self-reported lies. *Human Communication Research, 36*(1), 2–25.

Seuntjens, T. G., Zeelenberg, M., van de Ven, N., & Breugelmans, S. M. (2019). Greedy bastards: Testing the relationship between wanting more and unethical behavior. *Personality and Individual Differences, 138*, 147–156. https://doi.org/10.1016/j.paid.2018.09.027.

Shakespeare, W., (2019) All's well that ends well, In One Hour Shakespear, pp.19-83, Routledge.

Shalvi, S., Dana, J., Handgraaf, M. J. J., & de Dreu, C. K. W. (2011). Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes, 115*(2), 181–190. https://doi.org/10.1016/j.obhdp.2011.02.001.

Shalvi, S., & de Dreu, C. K. W. (2014). Oxytocin promotes group-serving dishonesty. *Proceedings of the National Academy of Sciences, 111*(15), 5503–5507. https://doi.org/10.1073/pnas.1400724111.

Shalvi, S., Eldar, O., & Bereby-Meyer, Y. (2012). Honesty requires time (and lack of justifications). *Psychological Science, 23*(10), 1264–1270. https://doi.org/10.1177/0956797612443835.

Shalvi, S., Gino, F., Barkan, R., & Ayal, S. (2015). Self-serving justifications: Doing wrong and feeling moral. *Current Directions in Psychological Science, 24*(2), 125–130. https://doi.org/10.1177/0963721414553264.

Shalvi, S., Handgraaf, M. J. J., & de Dreu, C. K. W. (2011). Ethical manoeuvring: Why people avoid both major and minor lies. *British Journal of Management, 22*(s1), https://doi.org/10.1111/j.1467-8551.2010.00709.x.

Shell, G. R. (2021). The Conscience Code: Lead with Your Values. Advance Your Career. *HarperCollins Leadership*.

Shirado, H., Hou, Y. T.-Y., & Jung, M. F. (2023). Stingy bots can improve human welfare in experimental sharing networks. *Scientific Reports, 13*(1), 17957. https://doi.org/10.1038/s41598-023-44883-0.

Shuster, S., Eyal, T., Ayal, S., & Moran, S. (2024). Proud to be dishonest: Emotional consequences of altruistic versus egoistic dishonesty. *Journal of Behavioral Decision Making, 37*(3), https://doi.org/10.1002/bdm.2386.

Skowronek, S. E. (2022). DENIAL: A conceptual framework to improve honesty nudges. *Current Opinion in Psychology, 48*, 101456. https://doi.org/10.1016/j.copsyc.2022.101456.

Snowden, A. W., Hancock, A. S., Buhusi, C. V., & Warren, C. M. (2022). Event-related correlates of evolving trust evaluations. *Social Neuroscience, 17*(2), 154–169. https://doi.org/10.1080/17470919.2022.2043935.

Sommerfeld, R. D., Krambeck, H.-J., Semmann, D., & Milinski, M. (2007). Gossip as an alternative for direct observation in games of indirect reciprocity. *Proceedings of the National Academy of Sciences, 104*(44), 17435–17440. https://doi.org/10.1073/pnas.0704598104.

Speer, S. P. H., Martinovici, A., Smidts, A., & Boksem, M. A. S. (2023). The acute effects of stress on dishonesty are moderated by individual differences in moral default. *Scientific Reports, 13*(1), 3984. https://doi.org/10.1038/s41598-023-31056-2.

Speer, S. P. H., Smidts, A., & Boksem, M. A. S. (2020). Cognitive control increases honesty in cheaters but cheating in those who are honest. *Proceedings of the National Academy of Sciences, 117*(32), 19080–19091. https://doi.org/10.1073/pnas.2003480117.

Speer, S. P., Smidts, A., & Boksem, M. A. S. (2021). Cognitive control promotes either honesty or dishonesty, depending on one's moral default. *The Journal of Neuroscience, 41*(42), 8815–8825. https://doi.org/10.1523/JNEUROSCI.0666-21.2021.

Speer, S. P. H., Smidts, A., & Boksem, M. A. S. (2022a). Cognitive control and dishonesty. *Trends in Cognitive Sciences, 26*(9), 796–808. https://doi.org/10.1016/j.tics.2022.06.005.

Speer, S. P. H., Smidts, A., & Boksem, M. A. S. (2022b). Individual differences in (dis) honesty are represented in the brain's functional connectivity at rest. *Neuroimage, 246*, 118761. https://doi.org/10.1016/j.neuroimage.2021.118761.

Starke, C., Ventura, A., Bersch, C., Cha, M., de Vreese, C., Doebler, P., ... Köbis, N. (2024). Risks and protective measures for synthetic relationships. *Nature Human Behaviour, 8*(10), 1834–1836. https://doi.org/10.1038/s41562-024-02005-4.

Starke, C., Kieslich, K., Reichert, M., & Köbis, N., (2023). Algorithms against corruption: A conjoint study on designing automated twitter posts to encourage collective action. https://doi.org/10.31235/osf.io/wf45t.

Stephens-Davidowitz, S. (2017). *Everybody lies: Big data, new data, and what the internet can tell us about who we really are*. Harperscollins Publishers.

Suchotzki, K., Verschuere, B., van Bockstaele, B., Ben-Shakhar, G., & Crombez, G. (2017). Lying takes time: A meta-analysis on reaction time measures of deception. *Psychological Bulletin, 143*(4), 428–453. https://doi.org/10.1037/bul0000087.

Sun, J., & Smillie, L. D. (2024). Why moral psychology needs personality psychology. *Journal of Personality, 92*(3), 653–665. https://doi.org/10.1111/jopy.12919.

Sutter, M. (2009). Deception through telling the truth?! Experimental evidence from individuals and teams. *The Economic Journal, 119*(534), 47–60. https://doi.org/10.1111/j.1468-0297.2008.02205.x.

Swann, W. B., & Buhrmester, M. D. (2015). Identity fusion. *Current Directions in Psychological Science, 24*, 52–57. https://doi.org/10.1177/0963721414551363.

Swann, W. B., Gómez, A., Huici, C., Morales, J. F., & Hixon, J. G. (2010). Identity fusion and self-sacrifice: Arousal as a catalyst of pro-group fighting, dying, and helping behavior. *Journal of Personality and Social Psychology, 99*, 824–841.

Swann, W. B., Gómez, Á., Seyle, D. C., Morales, J. F., & Huici, C. (2009). Identity fusion: The interplay of personal and social identities in extreme group behavior. *Journal of Personality and Social Psychology, 96*(5), 995–1011. https://doi.org/10.1037/a0013668.

Swann, W. B., Jetten, J., Gómez, Á., Whitehouse, H., & Bastian, B. (2012). When group membership gets personal: A theory of identity fusion. *Psychological Review, 119*(3), 441–456.

Talwar, V., & Lee, K. (2002). Development of lying to conceal a transgression: Children's control of expressive behaviour during verbal deception. *International Journal of Behavioral Development, 26*(5), 436–444. https://doi.org/10.1080/01650250143000373.

Talwar, V., & Lee, K. (2008). Social and cognitive correlates of children's lying behavior. *Child Development, 79*(4), 866–881. https://doi.org/10.1111/j.1467-8624.2008.01164.x.

Teodorescu, K., & Erev, I. (2014). Learned helplessness and learned prevalence. *Psychological Science, 25*(10), 1861–1869. https://doi.org/10.1177/0956797614543022.

Teodorescu, K., Plonsky, O., Ayal, S., & Barkan, R. (2021). Frequency of enforcement is more important than the severity of punishment in reducing violation behaviors. *Proceedings of the National Academy of Sciences, 118*(42), https://doi.org/10.1073/pnas.2108507118.

Testori, M., Hemelrijk, C. K., & Beersma, B. (2022). Gossip promotes cooperation only when it is pro-socially motivated. *Scientific Reports, 12*(1), 4790. https://doi.org/10.1038/s41598-022-08670-7.

Thaler, R. H., & Sunstein, C. R. (2021). *Nudge: The final edition*. Penguin.

Thielmann, I., Böhm, R., & Hilbig, B. E. (2021). Buying unethical loyalty: A behavioral paradigm and empirical test. *Social Psychological and Personality Science, 12*(3), 363–370. https://doi.org/10.1177/1948550620905218.

Thielmann, I., Hilbig, B. E., Klein, S. A., Seidl, A., & Heck, D. W. (2024). Cheating to benefit others? On the relation between honesty-humility and prosocial lies. *Journal of Personality, 92*(3), 870–882. https://doi.org/10.1111/jopy.12835.

Thielmann, I., Hilbig, B. E., Zettler, I., & Moshagen, M. (2017). On measuring the sixth basic personality dimension: A comparison between HEXACO honesty-humility and big six honesty-propriety. *Assessment, 24*(8), 1024–1036. https://doi.org/10.1177/1073191116638411.

Thielmann, I., & Hilbig, B. E. (2018). Daring dishonesty: On the role of sanctions for (un)ethical behavior. *Journal of Experimental Social Psychology, 79*, 71–77.

Thielmann, I., Hilbig, B. E., Schild, C., & Heck D. W. (2024). Cheat, cheat, repeat: On the consistency of dishonest behavior in structurally comparable situations. doi: 10.31234/osf.io/ntg6c.

Thornton, D., Gunningham, N. A., & Kagan, R. A. (2005). General deterrence and corporate environmental behavior. *Law & Policy, 27*(2), 262–288. https://doi.org/10.1111/j.1467-9930.2005.00200.x.

Tobsch, N., Sorcha, L., & Van Rooij, B. (2024). Can organizational culture predict major wrongdoing? A review of existing survey measures. *Draft Paper Under Review*.

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition, 27*(6), 813–833. https://doi.org/10.1521/soco.2009.27.6.813.

Tyler, T. R. (2006). *Why people obey the law*. Princeton University Press.

Urgesi, C., Aglioti, S. M., Skrap, M., & Fabbro, F. (2010). The spiritual brain: Selective cortical lesions modulate human self-transcendence. *Neuron, 65*(3), 309–319. https://doi.org/10.1016/j.neuron.2010.01.026.

Urgesi, C., Berlucchi, G., & Aglioti, S. M. (2004). Magnetic stimulation of extrastriate body area impairs visual processing of nonfacial body parts. *Current Biology, 14*(23), 2130–2134. https://doi.org/10.1016/j.cub.2004.11.031.

Urgesi, C., Calvo-Merino, B., Haggard, P., & Aglioti, S. M. (2007a). Transcranial magnetic stimulation reveals two cortical pathways for visual body processing. *Journal of Neuroscience, 27*(30), 8023–8030. https://doi.org/10.1523/JNEUROSCI.0789-07.2007.

Urgesi, C., Candidi, M., Ionta, S., & Aglioti, S. M. (2007b). Representation of body identity and body actions in extrastriate body area and ventral premotor cortex. *Nature Neuroscience,* (1), 30–31. https://doi.org/10.1038/nn1815.

Vabba, A., Porciello, G., Panasiti, M. S., & Aglioti, S. M. (2022). Interoceptive influences on the production of self-serving lies in reputation risk conditions. *International Journal Psychophysiology, 177*, 34–42. https://doi.org/10.1016/j.ijpsycho.2022.04.001.

van der Cruyssen, I., D'Hondt, J., Meijer, E., & Verschuere, B. (2019). Does honesty require time? Two preregistered replications of Experiment 2 of Shalvi, Eldar, and Bereby-Meyer (2012). https://doi.org/10.31219/osf.io/qbk6z.

van Lange, P. A. M., & Balliet, D. (2015). *Interdependence theory. APA handbook of personality and social psychology, Volume 3: Interpersonal relations*. American Psychological Association, 65–92. https://doi.org/10.1037/14344-003.

van Rooij, B. (2016). Weak enforcement, strong deterrence: Dialogues with chinese lawyers about tax evasion and compliance. *Law & Social Inquiry, 41*(02), 288–310. https://doi.org/10.1111/lsi.12136.

Verschuere, B., Köbis, N. C., Bereby-Meyer, Y., Rand, D., & Shalvi, S. (2018). Taxing the brain to uncover lying? Meta-analyzing the effect of imposing cognitive load on the reaction-time costs of lying. *Journal of Applied Research in Memory and Cognition,* 7(3), 462–469. https://doi.org/10.1016/j.jarmac.2018.04.005.

Verschuere, B., & Shalvi, S. (2014). The Truth comes naturally! Does it? *Journal of Language and Social Psychology, 33*(4), 417–423. https://doi.org/10.1177/0261927X14535394.

Vincent, L. C., & Kouchaki, M. (2016). Creative, rare, entitled, and dishonest: How commonality of creativity in one's group decreases an individual's entitlement and dishonesty. *Academy of Management Journal, 59*(4), 1451–1473. https://doi.org/10.5465/amj.2014.1109.

Vohs, K. D., Schmeichel, B. J., Lohmann, S., Gronau, Q. F., Finley, A. J., Ainsworth, S. E., ... Albarracín, D. (2021). A multisite preregistered paradigmatic test of the ego-depletion effect. *Psychological Science, 32*(10), 1566–1581. https://doi.org/10.1177/0956797621989733.

Von Hippel, W., & Trivers, R. (2011). The evolution and psychology of self-deception. *Behavioral and brain sciences, 34*(1), 1–16.

von Schenk, A., Klockmann, V., Bonnefon, J.-F., Rahwan, I., & Köbis, N. (2024). Lie detection algorithms disrupt the social dynamics of accusation behavior. *IScience, 27*(7), 110201. https://doi.org/10.1016/j.isci.2024.110201.

Vrij, A., Fisher, R., Mann, S., & Leal, S. (2008). A cognitive load approach to lie detection. *Journal of Investigative Psychology and Offender Profiling, 5*(1–2), 39–43. https://doi.org/10.1002/jip.82.

Walters, G. D., & Bolger, P. C. (2019). Procedural justice perceptions, legitimacy beliefs, and compliance with the law: A meta-analysis. *Journal of Experimental Criminology, 15*(3), 341–372. https://doi.org/10.1007/s11292-018-9338-2.

Weisburd, D., Maher, L., Sherman, L., Buerger, M., Cohn, E., & Petrosino, A. (2023). *Contrasting crime general and crime specific theory: The case of hot spots of crime. New directions in criminological theory*. Routledge, 45–70. https://doi.org/10.4324/9781003421139-3.

Weisel, O., & Shalvi, S. (2015). The collaborative roots of corruption. *Proceedings of the National Academy of Sciences, 112*(34), 10651–10656. https://doi.org/10.1073/pnas.1423035112.

Weisel, O., & Shalvi, S. (2022). Moral currencies: Explaining corrupt collaboration. *Current Opinion in Psychology, 44*, 270–274. https://doi.org/10.1016/j.copsyc.2021.08.034.

West, C., & Zhong, C.-B. (2015). Moral cleansing. *Current Opinion in Psychology, 6*, 221–225. https://doi.org/10.1016/j.copsyc.2015.09.022.

Wildschut, T., & Insko, C. A. (2007). Explanations of interindividual–intergroup discontinuity: A review of the evidence. *European review of social psychology, 18*(1), 175–211.

Williams, L. E., Huang, J. Y., & Bargh, J. A. (2009). The scaffolded mind: Higher mental processes are grounded in early experience of the physical world. *European Journal of Social Psychology, 39*(7), 1257–1267. https://doi.org/10.1002/ejsp.665.

Wilson, A., Stefanik, C., & Shank, D. B. (2022). How do people judge the immorality of artificial intelligence versus humans committing moral wrongs in real-world situations? *Computers in Human Behavior Reports, 8*, 100229. https://doi.org/10.1016/j.chbr.2022.100229.

Wiltermuth, S. S. (2011). Cheating more when the spoils are split. *Organizational Behavior and Human Decision Processes, 115*(2), 157–168. https://doi.org/10.1016/j.obhdp.2010.10.001.

Wu, J., Balliet, D., & van Lange, P. A. M. (2016). Reputation, gossip, and human cooperation. *Social and Personality Psychology Compass, 10*(6), 350–364. https://doi.org/10.1111/spc3.12255.

Xu, J., Yang, T., Zhuang, S., Li, H., & Lu, W. (2024). AI-based financial transaction monitoring and fraud prevention with behaviour prediction. *Applied and Computational Engineering, 77*, 218–224.

Yagil, D., & Medler-Liraz, H. (2014). Feel free, be yourself: Authentic leadership, emotional expression, and employee authenticity. *Journal of Leadership & Organizational Studies, 21*(1), 59–70.

Yin, L., Reuter, M., & Weber, B. (2016). Let the man choose what to do: Neural correlates of spontaneous lying and truth-telling. *Brain and Cognition, 102*, 13–25. https://doi.org/10.1016/j.bandc.2015.11.007.

Yin, L., & Weber, B. (2019). I lie, why don't you: Neural mechanisms of individual differences in self-serving lying. *Human Brain Mapping, 40*(4), 1101–1113. https://doi.org/10.1002/hbm.24432.

Yin, L., Zhong, S., Guo, X., & Li, Z. (2021). Functional connectivity between the caudate and medial prefrontal cortex reflects individual honesty variations in adults and children. *Neuroimage, 238*, 118268. https://doi.org/10.1016/j.neuroimage.2021.118268.

Yuthas, K., Rogers, R., & Dillard, J. F. (2002). Communicative Action and Corporate Annual Reports. *Journal of Business Ethics, 41*, 141–157.

Zanetti, C., & Butera, F. (2023). The chronology of collective cheating: A qualitative study of collective dishonesty in academic contexts. *Current Psychology, 42*(32), 27983–27997. https://doi.org/10.1007/s12144-022-03885-3.

Zanetti, C., & Butera, F. (2025). Detecting collecting cheating in academic contexts. *Educational Psychology, 45*(1), https://doi.org/10.1080/01443410.2025.2449976.

Zettler, I., Thielmann, I., Hilbig, B. E., & Moshagen, M. (2020). The nomological net of the HEXACO model of personality: A large-scale meta-analytic investigation. *Perspectives on Psychological Science, 15*(3), 723–760. https://doi.org/10.1177/1745691619895036.

Zhang, H., Shi, Y., Zhou, Z. E., Ma, H., & Tang, H. (2020). Good people do bad things: How anxiety promotes unethical behavior through intuitive and automatic processing. *Current Psychology, 39*(2), 720–728. https://doi.org/10.1007/s12144-018-9789-7.

Zhong, C. B., & Liljenquist, K. (2006). Washing away your sins: Threatened morality and physical cleansing. *Science (New York, N. Y.), 313*, 1451–1452. https://doi.org/10.1126/science.1130726.

Zhu, L., Jenkins, A. C., Set, E., Scabini, D., Knight, R. T., Chiu, P. H., ... Hsu, M. (2014). Damage to dorsolateral prefrontal cortex affects tradeoffs between honesty and self-interest. *Nature Neuroscience, 17*(10), 1319–1321. https://doi.org/10.1038/nn.3798.

Zickfeld, J. H., Karg, S. T. S., Engen, S. S., Gonzalez, A. S. R., Michael, J., & Mitkidis, P. (2024). Committed (dis)honesty: A systematic meta-analytic review of the divergent effects of social commitment to individuals or honesty oaths on dishonest behavior. *Psychological Bulletin, 150*(5), 586–620. https://doi.org/10.1037/bul0000429.

Zickfeld, J., Scigala, K., Elbaek, C. T., Michael, J., Tønnesen, M. T., Levy, G., ... Mitkidis, P. (2024). I solemnly swear I'm up to good: A megastudy investigating the effectiveness of honesty oaths on curbing dishonesty. *Nature Human Behavior*. https://doi.org/10.31234/osf.io/hctxe.

Zickfeld, J. H., Ścigała, K. A., Weiss, A., Michael, J., & Mitkidis, P. (2023). Commitment to honesty oaths decreases dishonesty, but commitment to another individual does not affect dishonesty. *Communications Psychology, 1*(1), 27. https://doi.org/10.1038/s44271-023-00028-7.

Zuo, X.-N., & Xing, X.-X. (2014). Test-retest reliabilities of resting-state FMRI measurements in human brain functional connectomics: A systems neuroscience perspective. *Neuroscience & Biobehavioral Reviews, 45*, 100–118. https://doi.org/10.1016/j.neubiorev.2014.05.009.