Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Psychology

ELSEVIER

Review

When does moral engagement risk triggering a hypocrite penalty?

Jillian Jordan[1,a] and Roseanna Sommers[2,a]

Abstract

Society suffers when people stay silent on moral issues. Yet people who engage morally may appear *hypocritical* if they behave imperfectly themselves. Research reveals that hypocrites can—but do not always—trigger a "hypocrisy penalty," whereby they are evaluated as more immoral than ordinary (non-hypocritical) wrongdoers. This pattern reflects that moral engagement can confer reputational *benefits*, but can also carry reputational *costs* when paired with inconsistent moral conduct. We discuss mechanisms underlying these costs and benefits, illuminating when hypocrisy is (and is not) evaluated negatively. Our review highlights the role that dishonesty and other factors play in engendering disdain for hypocrites, and offers suggestions for how, in a world where nobody is perfect, people can engage morally without generating backlash.

Addresses

[1] Harvard Business School, Soldiers Field Rd. Boston, MA 02163, USA
[2] University of Michigan Law School, 701 South State St. Ann Arbor, MI 48109, USA

Corresponding author: Jordan, Jillian (jijordan@hbs.edu)
[a] These authors contributed equally.

Current Opinion in Psychology 2022, **47**:101404

This review comes from a themed issue on **Honesty and Deception**

Edited by **Maurice E. Schweitzer** and **Emma Levine**

For a complete overview see the Issue and the Editorial

Available online 24 June 2022

https://doi.org/10.1016/j.copsyc.2022.101404

Keywords

Hypocrisy, Deception, Honesty, Morality, Signaling, Reputation.

Introduction

On March 8, 2022—International Women's Day—companies across the globe released statements on social media celebrating their female employees and touting their firms' commitment to gender equality. For example, one company posted on Twitter, "At [our company] we are proud to celebrate women in senior leadership, who have found great job prospects across the UK." A user named @paygapapp instantly replied to the tweet with a message of its own: "In this organization, women's median hourly pay is 8.8% lower than men's" [1]. The @paygapapp account, a bot, was programmed to identify companies tweeting about International Women's Day and respond automatically with a quote-tweet sharing the salary disparities, gleaned from publicly available data, between male and female employees at the organization [2]. The app's creators built it to "enable the public to hold companies to account over the words of 'empowerment', 'inspiration', and 'celebration' they tweet on International Women's Day."

This kind of blowback isn't unique to companies posting empty platitudes. Firms pursuing impactful Corporate Social Responsibility activities risk backlash if they couch their efforts in moral language, as opposed to business strategy [3]. Even spearheading a charitable initiative can backfire. When Meghan Markle celebrated her fortieth birthday by asking forty of her friends to each donate forty minutes of their time to mentoring women re-entering the workforce, she faced severe criticism. One commentator said, "I don't want to see her lecturing young mums having to go back to work from inside her $11 million LA mansion."

These examples illustrate how *moral engagement* (e.g., expressing moral opinions, urging others to donate their time or money, or calling out bad behavior) can feel perilous. While many people care about moral issues and wish to engage with them—an outcome that is productive for society—most of us do not have a perfect track record of moral conduct. Therefore, engaging morally might seem like stepping out onto a tightrope: one misstep in your personal behavior, and you risk being brought down by your *hypocrisy*. Research shows that hypocrites can be particularly disliked, and judged as morally worse than ordinary (i.e., non-hypocritical) wrongdoers who never engaged morally in the first place [3–7]. In this review, we explore when moral engagement does—and does not—risk triggering this *hypocrisy penalty*.

It's surprisingly easy to be judged a hypocrite

It may seem obvious that some forms of moral engagement, like calling out bad behavior, set you up to appear hypocritical, should you behave badly yourself. But evidence suggests that even without publicly criticizing others, it is surprisingly easy to be judged a hypocrite.

Indeed, research reveals that *many* forms of moral engagement can be deemed hypocritical when paired with less-than-perfect moral conduct [3,7–17]. For example, one study found that 43% of subjects judge an individual to be a hypocrite for volunteering at a church bake sale despite also sometimes watching adult films, and 65% of subjects say it's hypocritical to privately believe that illegal drug use is wrong but nonetheless smoke marijuana [17]. Another study showed that privately donating money to anti-smoking causes can appear hypocritical, if the donation comes from a tobacco company executive [16]. And leaders who take moral stances can be deemed hypocritical if they later change their minds [7].

Findings like these might suggest that any moral engagement poses an inherent liability, because any perceived inconsistency between your moral engagement and your personal behavior opens you to criticism. Perhaps, then, it is safer to inoculate yourself from charges of hypocrisy by choosing moral apathy.

Yet not all hypocrites are judged negatively

Just because moral engagement can put you at risk for being seen as a hypocrite, however, does not necessarily mean that you risk being judged *negatively* for your hypocrisy. This distinction, between *hypocrisy* and *negative moral evaluation*, has been underappreciated in the psychological literature on hypocrisy. For instance, it has been argued that research on hypocrisy is important "because individuals and organizations suffer consequences when they are perceived as hypocritical" [16] and that interpreting a target's inconsistent behavior as hypocritical necessarily and directly leads to evaluating the target negatively [18].

Yet in some cases, a target who engages morally and transgresses is rated as more *hypocritical* than a target who simply transgresses, but is judged to be no more [4,7,19,20], or even less [16,21], *immoral*. For example, Huppert and colleagues assigned subjects to evaluate a politician who lied about the source of his campaign funds. They found that if the politician took the strong stance that "it is never okay to lie," he was rated as more hypocritical—but *less* immoral—than his counterpart who stated that it is sometimes okay to lie. Subjects were also more inclined to vote for the hypocrite.

Thus, not all acts that are interpreted as hypocritical trigger negative moral evaluation. It seems that hypocritical moral engagement can confer both reputational benefits *and* costs, and thus can result in net reputational gains *or* losses (Fig. 1). So, if we want to encourage productive moral engagement in a world where nobody is perfect, a key question is how people can maximize the reputational *benefits* that flow from their engagement while minimizing the reputational *costs*.

Reputational benefits of moral engagement

A large body of research suggests that, in isolation (i.e., when not paired with a moral transgression), moral engagement tends to be perceived positively. For example, individuals' reputations are bolstered when they express moral values [22–24], behave prosocially [25–32], and condemn and punish wrongdoers [19,27,33–39] (Correspondingly, their moral reputations can suffer when they choose to "stay out of it" on moral issues [40]). Moreover, the aforementioned evidence that hypocrites (such as the dishonest politician who takes a strong stance against lying) can sometimes be evaluated *more positively* than non-hypocritical transgressors (such as a dishonest politician who does not take such a stance) suggests that moral engagement can continue to confer reputational benefits even when paired with a moral transgression.

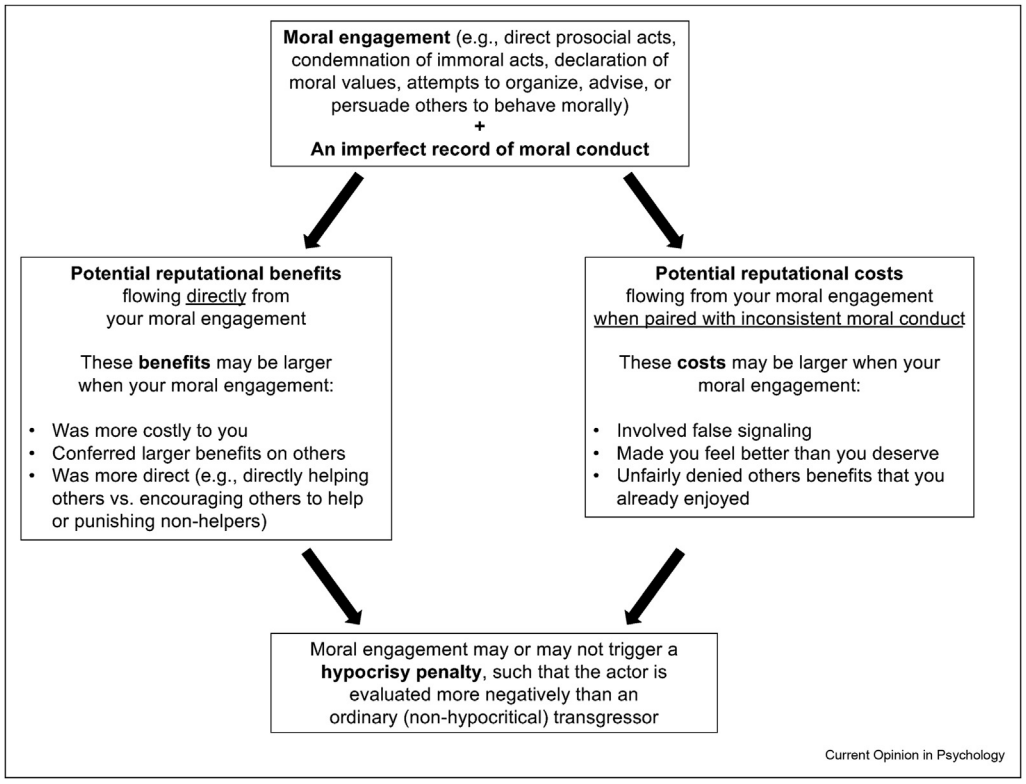
Furthermore, the literature highlights several ways in which one might maximize the reputational benefits that flow from moral engagement. In particular, moral engagement tends to be perceived especially positively when it is (i) more costly to the actor (e.g., because it requires a greater investment of time, effort, or resources) [29,35,37,41–45], (ii) more beneficial to others [43–47] (although observers may be relatively less sensitive to benefits achieved than costs incurred [29,43–45,48]), and (iii) more direct (e.g., cooperating tends to confer larger reputational benefits than punishing non-cooperation in others [26,27]).

Reputational costs of hypocritical moral engagement

Yet moral engagement does not *always* remain a net reputational positive when paired with a moral transgression that creates perceived inconsistency. For example, in isolation, condemning others' transgressions (vs. staying silent) can burnish the condemner's reputation, making her appear more moral. In addition, those who condemn (e.g., telling others "I think cheating is morally wrong") are judged as even more moral than individuals who *directly* state that they do not engage in the relevant behavior (e.g., "I never cheat.") [4]. It seems that moral condemnation can generate especially strong reputational benefits. However, when paired with the revelation that the speaker privately commits the relevant transgression, the same condemnation can make the speaker look *worse* than she would have had she stayed silent. Furthermore, a cheater who has (hypocritically) condemned cheating is judged as less moral than a cheater who has (falsely) stated that she does not cheat [4], demonstrating that hypocrites can be disliked even more than direct liars.

Thus, moral engagement that is perceived as a strong *positive* in isolation can carry reputational *costs* when coupled with bad behavior. But why? This question, in

Figure 1



A causal model of the reputational consequences of moral engagement, when paired with moral conduct that is perceived as inconsistent. On the one hand, moral engagement can have direct reputational *benefits*. On the other hand, moral engagement that is paired with an imperfect moral track record (creating perceived inconsistency) can also carry reputational *costs*. When the costs outweigh the benefits, hypocritical moral engagement is liable to trigger a "hypocrisy penalty," whereby the actor is evaluated more negatively than a non-hypocritical transgressor who avoided moral engagement. See Section 4 for further discussion of the reputational *benefits* that can flow from moral engagement, and Sections 5.1-5.3 for further discussion of the reputational *costs* (as well as examples of situations that may trigger these costs).

our view, represents the core puzzle of hypocrisy. Given that moral engagement is normally seen as *virtuous*, why should it ever *exacerbate* (rather than mitigate) our negative evaluations of transgressors? Here, we (i) discuss three potential mechanisms through which hypocritical moral engagement may create reputational costs, and thus give rise to the hypocrisy penalty (in Sections 5.1-5.3), and (ii) review evidence hinting at mechanisms that do *not* seem sufficient to trigger a hypocrisy penalty (in Section 5.4).

False signaling

Hypocritical moral engagement may be judged negatively when it is misleading to others [4,10,49,50]. In general, moral engagement can convey information about the actor's personal conduct. For example, a target who criticizes wrongdoing is judged to be less likely to commit similar transgressions herself, as compared to someone who stays silent [4]. Therefore, those who privately transgress after engaging morally may be *false signalers*, if their moral engagement implies that they behave more morally than they actually do.

Furthermore, evidence suggests that false signaling can contribute to the reputational costs of hypocritical moral engagement. In one study, "traditional" hypocrites (who sent false signals by publicly condemning acts that they privately engaged in) suffered a hypocrisy penalty. However, "honest" hypocrites—who *avoided* false signaling, by publicly admitting to the transgressions they condemned (e.g., "I think it's morally wrong when people try to get out of jury duty, but I sometimes do it anyway")—escaped the hypocrisy penalty. These honest hypocrites were judged to be no less moral (but still much more hypocritical) than non-hypocritical transgressors [4] (Fig. 2).

Might this finding instead reflect that honest hypocrites *are* penalized for their hypocrisy, but are also rewarded for their willingness to openly confess their transgressions? To rule out this possibility, participants also evaluated hypocrites who engaged in false signaling (e.g., by condemning single-sided printing, and then subsequently printing single-sided) but openly admitted to *unrelated* transgressions (e.g., downloading

music illegally). These hypocrites *were* evaluated more negatively than non-hypocritical transgressors, suggesting that openly confessing to negative behavior will protect hypocrites from negative evaluation only when the confession serves to negate the false signal [4]. Taken together, these findings provide strong evidence that false signaling can contribute to the reputational costs of hypocritical moral engagement.

Feeling more moral than is merited

Hypocritical moral engagement may also be judged negatively when the actor is perceived, because of their moral engagement, to *feel more moral than is merited*. Effron and colleagues (2018) propose that a key ingredient of hypocrisy is "claim [ing] an undeserved moral benefit," where a moral benefit is defined as any social or psychological reward for virtuous behavior [18]. Consistent with this proposal, O'Connor and colleagues (2020) find that even when moral engagement does not involve false signaling, it can carry reputational costs, insofar as the target seems to enjoy the psychological benefit of feeling better than is warranted [16].

For example, a tobacco executive who secretly donates to anti-smoking causes is not false signaling, because he earns no reputational benefits from the portions of his behavior that are public. Yet O'Connor and colleagues find evidence that subjects see his anti-smoking donation as hypocritical, and infer that it makes him feel better than is merited (by alleviating his guilt about working in the tobacco industry). Consequently, subjects judge the executive more negatively than if he had donated to an unrelated cause (e.g., an anti-obesity charity) [16].

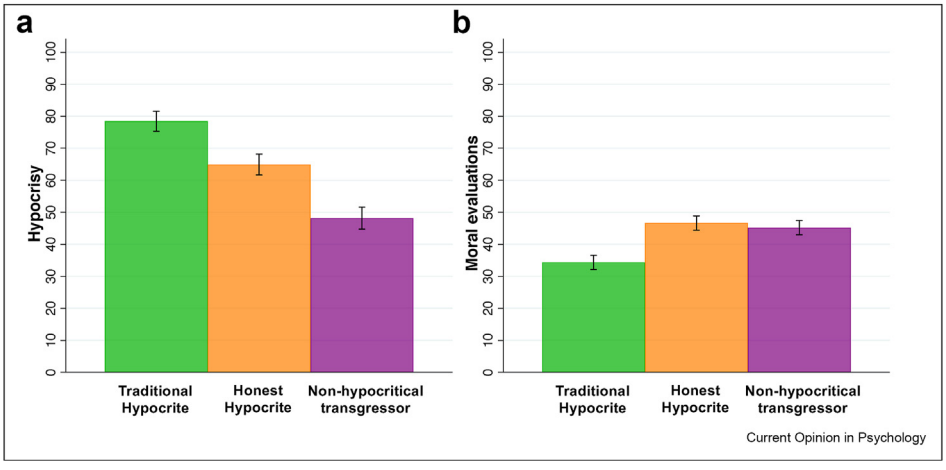
Pressuring others to behave virtuously despite enjoying the benefits of vice

Hypocritical moral engagement may also carry reputational costs when actors are seen as unfairly denying others benefits that they have personally enjoyed. For instance, consider a father who insists that his teenage son abstain from all alcohol and drugs despite having, by his own admission, enjoyed the thrills of living irresponsibly in his youth. While the father isn't necessarily falsely signaling or feeling better than he deserves, he might be judged negatively for unfairly demanding that his son, unlike him, live a self-denying adolescence. Consistent with this idea, previous research shows that actors who have benefited (vs. suffered) from past deeds that they go on to preach against are evaluated more negatively [51]. This pattern may reflect that people believe such actors do not have the right to influence others to be more virtuous once they have enjoyed the benefits of being less virtuous [18,51].

Mechanisms that do *not* seem sufficient to trigger a hypocrisy penalty

Prior research also hints at mechanisms that do *not* seem sufficient to trigger a hypocrisy penalty. Recall that "honest hypocrites" openly admit to engaging in the same transgressions they condemn (e.g., "I think it's morally wrong to use a lot of paper by printing

Figure 2



False signaling contributes to the reputational costs of hypocrisy. Displayed are data from Jordan et al., 2017, Study 4. Subjects in this study evaluated either (i) "traditional hypocrites," who sent false signals by publicly condemning immoral acts that they privately engaged in; (ii) "honest hypocrites," who condemned and committed the same immoral acts, but negated any false signaling by openly confessing to their transgressions; or (iii) "non-hypocritical transgressors," who committed the same immoral acts, but did not condemn them. **(a)** Both traditional hypocrites and honest hypocrites were rated as more *hypocritical* than non-hypocritical transgressors **(b)** Yet only *traditional* hypocrites, who engaged in false signaling, suffered a "hypocrisy penalty," earning more negative moral evaluations than non-hypocritical transgressors. These data demonstrate that not all hypocrisy is evaluated negatively, and suggest that false signaling is one mechanism through which hypocrisy can have reputational costs (and thus trigger a hypocrisy penalty). In this way, the results highlight the importance of avoiding moral engagement that appears dishonest.

documents single-sided, but sometimes I do it anyway"). As described previously, honest hypocrites escape the hypocrisy penalty: they are rated as no less moral than non-hypocritical transgressors (e.g., paper-wasters who do *not* declare paper-wasting immoral) [4,52].

Yet while honest hypocrites are not *penalized* for their hypocrisy, their hypocritical moral engagement still shapes how they are perceived. Evidence suggests that, relative to non-hypocritical transgressors, honest hypocrites are more likely to be seen as having violated a genuinely held moral value (or as having committed a "personal moral failing" [52]) by transgressing. In a recent study, participants judged that, by transgressing, honest hypocrites were more likely than non-hypocrites to have acted inconsistently with their values and to have intentionally done something they knew to be wrong. They also rated honest hypocrites as more weak-willed than non-hypocritical transgressors [52].

Because one might expect these inferences to reflect negatively on the moral character of honest hypocrites, it is notable that the study found no hypocrisy penalty for honest hypocrites. Therefore, this pattern of results illuminates several mechanisms that do *not* seem to reliably inflict reputational costs that are sufficient to trigger a hypocrisy penalty: seeming to suffer a weakness of will, appearing inconsistent with your moral values, and being perceived as *intentionally* committing a known wrongdoing.

How does the nature of the transgression shape evaluations of hypocrisy?

While this review has focused on properties of *moral engagement* that shape whether hypocrisy is evaluated negatively, properties of the actor's *moral transgression* might also affect judgments. For example, we might expect *highly severe* transgressions to create a "floor effect," whereby actors are judged intensely negatively regardless of whether or not they engage morally, thereby diminishing the hypocrisy penalty. Consistent with this proposal, in one study [20], a man who physically assaulted his girlfriend was seen as much more hypocritical when he was the spokesperson for an organization called "Stop the Violence" (vs. "Stop the Looting"); however, judgments relevant to his perceived moral character were comparable (and quite negative) regardless of which organization he served. This result suggests that the severity of his transgression may have overwhelmed any moral penalty he might have incurred specifically from his hypocrisy. On the other hand, more severe or intentional transgressions could plausibly cause an actor's moral engagement to seem less sincere and more calculating—exacerbating the costs associated with false signaling and thus *enhancing* the hypocrisy penalty. Future research might investigate these divergent possibilities, and more generally explore how the nature of one's transgression may influence whether hypocrisy is penalized.

Conclusion: how to walk the moral tightrope

To summarize, hypocrites can—but do not always—incur a "hypocrisy penalty," whereby they are evaluated more negatively than they would have been absent engaging. As this review has suggested, when observers scrutinize hypocritical moral engagement, they seem to ask at least three questions. First, does the actor signal to others, through his engagement, that he behaves more morally than he actually does? Second, does the actor, by virtue of his engagement, see himself as more moral than he really is? And third, is the actor's engagement preventing others from reaping benefits that he has already enjoyed? Evidence suggests that hypocritical moral engagement is more likely to carry reputational costs when the answer to these questions is "yes." At the same time, observers do *not* seem to reliably impose a hypocrisy penalty just because the transgressions of hypocrites constitute personal moral failings—even as these failings convey weakness of will, highlight inconsistency with the actor's personal values, and reveal that the actor has knowingly done something that she believes to be wrong.

In a world where nobody is perfect, then, how can one engage morally while limiting the risk of subsequently being judged negatively as a hypocrite? We suggest that the answer comes down to two key factors: maximizing the reputational benefits that flow directly from one's moral engagement, and minimizing the reputational costs that flow from the combination of one's engagement and imperfect track record. While more research is needed, here we draw on the mechanisms we have reviewed to highlight four suggestions for those seeking to walk the moral tightrope.

1. Do more good. Engaging in *costly* prosocial behavior that confers *meaningful* benefits upon others is likely to be perceived more positively than engaging in low-cost acts of "slacktivism," encouraging *others* to get involved without acting yourself, or simply condemning the wrongdoing of others. Therefore, acts of moral engagement that are direct, costly, and impactful may be less prone to turn into reputational liabilities, in the event that you are revealed to have a less-than-perfect track record.
2. Avoid false signaling—i.e., hiding your bad behavior while broadcasting your virtue. One way to avoid false signaling is to keep your moral engagement to yourself. But many forms of moral engagement are inherently social, and their positive impact depends on their being public. A more promising strategy, then, may be to engage publicly but readily acknowledge your inconsistency, rather than trying to wring maximal reputational benefits from your

engagement. Like the "honest hypocrite," pair your moral engagement with an admission that you yourself do not have a perfect record. And critically, when you engage morally in one domain (e.g., racial justice), you must own up to your shortcomings *in that same domain*; confessing to unrelated transgressions (e.g., an imperfect track record on environmental issues) will not suffice to avoid appearing dishonest.
3. Avoid giving off the impression that you think more highly of yourself than your mixed track record warrants, or that your moral engagement has served to cleanse your guilty conscience. Instead, attempt to convey that your self-image is shaped both by your positive *and* negative deeds.
4. If you are going to urge others to follow moral rules that you have personally violated, emphasize the costs that you suffered as a result of your transgressions. Avoid appearing to have unfairly reaped the benefits of flouting moral rules while denying those benefits to others.

Competing interest statement

None declared.

Acknowledgements

We thank Emmet Sullivan, Charlotte Gemperle, and Warda Yousef for helpful research assistance.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Mellor S: **International women's day: Twitter bot shames companies with gender pay gap figures**. *Fortune* 2022. Fortune, https://fortune.com/2022/03/09/twitter-paygapapp-bot-international-womens-day-gender-pay-gap-figure-shaming/. Accessed 30 April 2022; 2022.

2. Quito A: **A genius Twitter bot is calling out companies that post platitudes for International Women's Day**. *Quartz* 2022. https://qz.com/work/2139235/the-creators-of-gender-pay-gap-twitter-bot-explain-their-goals/. Accessed 30 April 2022; 2022.

3. ** Hafenbrädl S, Waeger D: **The business case for CSR: a trump card against hypocrisy?** *J Bus Res* 2021, **129**:838–848.
Reveals that businesses risk blowback when justifying their CSR activities with moral arguments. In contrast, providing a "business case" for CSR can help inoculate firms against charges of hypocrite.

4. Jordan JJ, Sommers R, Bloom P, Rand DG: **Why do we hate hypocrites? Evidence for a theory of false signaling**. *Psychol Sci* 2017, **28**:356–368.

5. Dong M, van Prooijen J-W, van Lange PA: **Calculating Hypocrites Effect: moral judgments of word-deed contradictory transgressions depend on targets' competence**. *Journal of Theoretical Social Psychology* 2021, **5**:489–501.

6. Dong M, van Prooijen J-W, Wu S, van Lange PA: **Culture, status, and hypocrisy: high-status people who don't practice what they preach are viewed as worse in the United States than China**. *Soc Psychol Personal Sci* 2022, **13**:60–69.

7. Kreps TA, Laurin K, Merritt AC: **Hypocritical flip-flop, or courageous evolution? When leaders change their moral minds**. *J Pers Soc Psychol* 2017, **113**:730.

8. Cha SE, Edmondson AC: **When values backfire: leadership, attribution, and disenchantment in a values-driven organization**. *Leader Q* 2006, **17**:57–78.

9. Effron DA, Lucas BJ, O'Connor K: **Hypocrisy by association: when organizational membership increases condemnation for wrongdoing**. *Organ Behav Hum Decis Process* 2015, **130**:147–159.

10. Graham J, Meindl P, Koleva S, Iyer R, Johnson KM: **When values and behavior conflict: moral pluralism and intrapersonal moral hypocrisy**. *Social and Personality Psychology Compass* 2015, **9**:158–170.

11. Hale Jr WJ, Pillow DR: **Asymmetries in perceptions of self and others' hypocrisy: rethinking the meaning and perception of the construct**. *Eur J Soc Psychol* 2015, **45**:88–98.

12. * Silver I, Newman G, Small DA: **Inauthenticity aversion: moral reactance toward tainted actors, actions, and objects**. *Consumer Psychology Review* 2021, **4**:70–82.
Provides a review of the concept of "inauthenticity aversion", and highlights that hypocritical moral engagement can appear inauthentic—a point that aligns with the proposal that false signaling can contribute to the reputational costs of hypocrite.

13. Wagner T, Lutz RJ, Weitz BA: **Corporate hypocrisy: overcoming the threat of inconsistent corporate social responsibility perceptions**. *J Market* 2009, **73**:77–91.

14. Barden J, Rucker DD, Petty RE: **"Saying one thing and doing another": examining the impact of event order on hypocrisy judgments of others**. *Pers Soc Psychol Bull* 2005, **31**:1463–1474.

15. Effron DA, Markus HR, Jackman LM, Muramoto Y, Muluk H: **Hypocrisy and culture: failing to practice what you preach receives harsher interpersonal reactions in independent (vs. interdependent) cultures**. *J Exp Soc Psychol* 2018, **76**:371–384.

16. ** O'Connor K, Effron DA, Lucas BJ: **Moral cleansing as hypocrisy: when private acts of charity make you feel better than you deserve**. *J Pers Soc Psychol* 2020, **119**:540.
Highlights that even in the absence of false signaling, hypocritical moral engagement can have reputational costs when the actor is perceived to feel more virtuous than is deserved.

17. Alicke M, Gordon E, Rose D: **Hypocrisy: what counts?** *Phil Psychol* 2013, **26**:673–701.

18. Effron DA, O'Connor K, Leroy H, Lucas BJ: **From inconsistency to hypocrisy: when does "saying one thing but doing another" invite condemnation?** *Res Organ Behav* 2018, **38**:61–75.

19. Kennedy JA, Schweitzer ME: **Building trust by tearing others down: when accusing others of unethical behavior engenders trust**. *Organ Behav Hum Decis Process* 2018, **149**:111–128.

20. Laurent SM, Clark BA, Walker S, Wiseman KD: **Punishing hypocrisy: the roles of hypocrisy and moral emotions in deciding culpability and punishment of criminal and civil moral transgressors**. *Cognit Emot* 2014, **28**:59–83.

21. ** Huppert E, Herzog N, Levine E, Landy J: *Being dishonest about dishonesty: the social benefits of taking absolute (but hypocritical) moral stances*. 2022. Working Paper.
Provides a potent demonstration that hypocrite can sometimes be perceived positively, owing to the reputational benefits of moral engagement (in this case, taking an absolute stand against lying).

22. Zlatev JJ: **I may not agree with you, but I trust you: caring about social issues signals integrity**. *Psychol Sci* 2019, **30**(6):880–892.

23. Van Zant AB, Moore DA: **Leaders' use of moral justifications increases policy support**. *Psychol Sci* 2015, **26**:934–943.

24. Kreps TA, Monin B: **Core values versus common sense: consequentialist views appear less rooted in morality**. *Pers Soc Psychol Bull* 2014, **40**:1529–1542.

25. Barclay P, Willer R: **Partner choice creates competitive altruism in humans**. *Proc Biol Sci* 2007, **274**:749–753, https://doi.org/10.1098/Rspb.2006.0209.

26. Raihani NJ, Bshary R: **Third-party punishers are rewarded–but third-party helpers even more so**. *Evolution* 2015.

27. Jordan JJ, Hoffman M, Bloom P, Rand D: **Third-party punishment as a costly signal of trustworthiness**. *Nature* 2016, **530**: 473–476.

28. Boyd R, Richerson PJ: **The evolution of indirect reciprocity**. *Soc Network* 1989, **11**:213–236.

29. * Berman JZ, Silver I: **Prosocial behavior and reputation: when does doing good lead to looking good?** *Current Opinion in Psychology* 2022, **43**:102–107.
A helpful overview of the contexts in which moral engagement that takes the form of prosocial behavior is most likely to confer substantial reputational benefits.

30. Jordan JJ, Hoffman M, Nowak MA, Rand DG: **Uncalculating cooperation is used to signal of trustworthiness**. *Proc Natl Acad Sci USA* 2016, **113**:8658–8663.

31. Barasch A, Levine EE, Berman JZ, Small DA: **Selfish or selfless? On the signal value of emotion in altruistic behavior**. *J Pers Soc Psychol* 2014, **107**(3):393–413.

32. Griskevicius V, Tybur JM, Van den Bergh B: **Going green to be seen: status, reputation, and conspicuous conservation**. *J Pers Soc Psychol* 2010, **98**:392.

33. Barclay P: **Reputational benefits for altruistic punishment**. *Evol Hum Behav* 2006, **27**:325–344, https://doi.org/10.1016/J.Evolhumbehav.2006.01.003.

34. Jordan JJ, Rand DG: **Signaling when nobody is watching: a reputation heuristics account of outrage and punishment in one-shot anonymous interactions**. *J Pers Soc Psychol* 2019.

35. Jordan JJ, Rand D: **Third-party punishment as a costly signal of high continuation probabilities in repeated games**. *J Theor Biol* 2017, **421**:189–202.

36. Raihani NJ, Bshary R: **The reputation of punishers**. *Trends Ecol Evol* 2015, **30**(2):98–103.

37. Nelissen R: **The price you pay: cost-dependent reputation effects of altruistic punishment**. *Evol Hum Behav* 2008, **29**: 242–248, https://doi.org/10.1016/J.Evothumbehav.2008.01.001.

38. Horita Y: **Punishers may be chosen as providers but not as recipients**. *Letters on Evolutionary Behavioral Science* 2010, **1**:6–9.

39. Hok H, Martin A, Trail Z, Shaw A: **When children treat condemnation as a signal: the costs and benefits of condemnation**. *Child Dev* 2019, **91**:1439–1455, https://doi.org/10.1111/cdev.13323.

40. ** Silver I, Shaw A: **When and why "staying out of it" backfires in moral and political disagreements**. *J Exp Psychol Gen* 2022.
Supports the proposal that moral engagement can have reputational benefits by highlighting the reputational costs of staying silent on moral issues.

41. Gintis H, Smith EA, Bowles S: **Costly signaling and cooperation**. *J Theor Biol* 2001, **213**:103–119.

42. Smith EA, Bliege Bird R: *Costly signaling and cooperative behavior, moral sentiments and material interests: the foundations of cooperation in economic life*. 2005:115–148.

43. Johnson S: *Dimensions of altruism: do evaluations of prosocial behavior track social good or personal sacrifice?*. 2018. Available at SSRN 3277444.

44. Erlandsson A, Wingren M, Andersson PA: **Type and amount of help as predictors for impression of helpers**. *PLoS One* 2020, **15**, e0243808.

45. Yudkin DA, Prosser A, Crockett MJ: **Actions speak louder than outcomes in judgments of prosocial behavior**. *Emotion* 2019, **19**:1138.

46. Delton AW, Robertson TE: **How the mind makes welfare tradeoffs: evolution, computation, and emotion**. *Current Opinion in Psychology* 2016, **7**:12–16.

47. Nemirow J: *The sacrifice halo: do we esteem altruists for the sacrifices they make or the benefits they deliver?*. Doctoral Dissertation. Harvard University; 2022.

48. Burum B, Nowak MA, Hoffman M: **An evolutionary explanation for ineffective altruism**. *Nat Human Behav* 2020, **4**: 1245–1257.

49. Monin B, Merritt A: *Moral hypocrisy, moral inconsistency, and the struggle for moral integrity*. 2012.

50. Batson CD, Thompson ER, Chen H: **Moral hypocrisy: addressing some alternatives**. *J Pers Soc Psychol* 2002, **83**: 330.

51. Effron DA, Miller DT: **Do as I say, not as I've done: suffering for a misdeed reduces the hypocrisy of advising others against it**. *Organ Behav Hum Decis Process* 2015, **131**: 16–32.

52. Jordan J, Sommers R: *False Signaling and Personal Moral Failings: two distinct pathways to hypocrisy with unequal moral weight*. 2020. Working Paper.