

RESEARCH ARTICLE

Hypocrisy judgements are affected by target attitude strength and attitude moralization

Thomas Ian Vaughan-Johnston 

School of Psychology, Cardiff University,
Cardiff, UK

Correspondence

Thomas Ian Vaughan-Johnston, School of
Psychology, Cardiff University, 70 Park Pl,
Cardiff, CF10 3AT, UK.
Email: thomasvaughanjohnston@gmail.com

Abstract

Researchers and philosophers have debated what leads people to judge others as being hypocritical. Some research has shown that perceivers consider targets to be more hypocritical when those targets contradict attitudes that are strongly (e.g., moralized and/or certain) rather than weakly held by the target. In the present work, I attempt to advance this research in several respects. First, I integrate these findings with research on the dimensions of attitude strength (i.e., commitment, embeddedness) to provide a more structured analysis of these claims. I show that characterizing a target's views as embedded *and* committed has many of the same hypocrisy-related effects as labelling those views as moral, and affect (negative) evaluations of targets through similar mechanisms. However, in Experiment 3, I show that moral attitudes are, nonetheless, perceived as distinct from classic strength dimensions in one crucial respect: the presumption that the target would impose them on other people. Furthermore, whereas judgements of hypocrisy relating to embedded/committed attitudes can be mitigated when perceivers engage in situational attribution, perceivers rendering judgements of hypocrisy relating to moral attitudes resist situational counter-explanations.

KEYWORDS

attitude strength, attitude-behaviour consistency, hypocrisy, moral judgement, morality, person perception

1 | INTRODUCTION

People often behave inconsistently from their attitudes, a principle that has been acknowledged in social psychology for almost a century ago (LaPiere, 1934; also see, e.g., Fazio & Zanna, 1981; Gross & Niman, 1975; Zanna et al., 1980). However, the frequency of such inconsistency in everyday life does not stop people from judging others, often very harshly, when their behaviours deviate from their expressed

attitudes.¹ A growing literature examines the conditions under which people (henceforth, 'perceivers') evaluate such actors (henceforth, 'targets') as hypocritical. Scholars have long noted that perceivers'

¹ Laurent and Clark (2019) also found modest hypocrisy judgements given attitude-attitude and behaviour-behaviour inconsistencies, but attitude-behaviour inconsistency was most strongly judged as hypocritical. Because my work focuses on perceivers' beliefs about target hypocrisy, I exclusively examined attitude-behaviour inconsistencies to capture hypocrisy.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2023 The Authors. *European Journal of Social Psychology* published by John Wiley & Sons Ltd.

labelling of hypocrisy deviates, often markedly, from philosophical conceptions of hypocrisy (e.g., Alicke et al., 2013). Nonetheless, what perceivers label 'hypocritical' matters (i.e., above and beyond strict philosophical definitions) because people render harsh social judgements towards hypocrites (Jordan et al., 2017; Teeny et al., 2023).

Aside from its obvious implications for the person perception literature, hypocrisy research can be viewed as drawing from and also having interesting implications for the attitudes literature. For example, some hypocrisy research draws from research on attitude strength (e.g., Krosnick et al., 1993) to examine how targets who express strong (vs. weak) attitudes may be judged as more hypocritical by perceivers (e.g., Laurent & Clark, 2019; Teeny et al., 2023). However, the attitude strength literature has developed some conceptual insights that have not been integrated into hypocrisy research, as I explore here. Hypocrisy research also draws from the study of moral attitudes as strong attitudes (Brandt & Wetherell, 2012; Luttrell et al., 2016, 2019; Mueller & Skitka, 2018; Skitka et al., 2005) to examine whether a target expressing a moral (vs. non-moral, such as an opinion held for practical reasons) attitude also may be judged as more hypocritical by perceivers when they behave inconsistently from it (e.g., Jordan et al., 2017; Kreps et al., 2017; Teeny et al., 2023). Yet the attitude scholars have raised theoretical claims about how the strength versus the moralization of attitudes may be similar and different, with a common claim being that moral attitudes are *not* simply strongly held attitudes (e.g., Brandt & Wetherell, 2012; Skitka et al., 2005). The present work examines whether these claims about moral attitudes have implications for hypocrisy research. Interestingly, increasing several even non-moral properties of the targets' expressed attitude strength may increase perceiver hypocrisy judgements despite holding the targets' behaviour remaining unchanged.

1.1 | Morality as 'just another' attitude strength dimension

Hypocrisy involving moral and strong attitudes may be similar. Teeny et al. (2023) found that hypocritical targets were appraised similarly, regardless of whether they contradicted moral or certain attitudes. Higher (vs. lower) attitude morality and certainty led to harsher hypocrisy judgements when targets contradicted these attitudes because perceivers were more surprised by the contradictory behaviour. Reducing surprise at acting against strong attitudes also lowered perceptions of hypocrisy.

However, the attitude strength literature has developed some conceptual distinctions (i.e., beyond low vs. high-attitude certainty) that may clarify how perceivers' evaluate hypocrites. Specifically, research has suggested that a wide range of attitude strength features (e.g., knowledge, importance, ambivalence, etc.; see Krosnick et al., 1993; Krosnick & Petty, 1995) may be underpinned by two underlying dimensions (Pomerantz et al., 1995; also see Philipp-Muller et al., 2020).

Commitment is associated with (low) perceived likelihood of changing their mind, extreme attitude positions and certainty that one's attitude position is right. In contrast, *embeddedness* represents how self-central an attitude is, and how important and value-expressive that attitude is for one.²

Moral attitudes' strength may be related to greater attitude commitment and/or embeddedness. Indeed, some past work shows moral attitudes being held with greater commitment in the sense of extremity (Brandt & Wetherell, 2012; Skitka et al., 2005) and certainty (Brandt & Wetherell, 2012); in other work, moral attitudes were shown to factor along with an embeddedness more than with a commitment factor (Philipp-Muller et al., 2020). Similarly, perceivers may treat targets' moral (vs. practical) attitudes as special (such as judging moral-attitude inconsistency more harshly; Teeny et al., 2023) because moral attitudes are committed, embedded attitudes. This is because acting against one's committed, embedded attitudes is both surprising and may reflect one's inauthenticity.

First, a target acting against their committed, embedded attitudes is probably surprising to perceivers – and Teeny et al. (2023) showed that surprise escalates hypocrisy judgements. A committed attitude is one that the attitudeholder expresses is unlikely to change; hence, it is surprising that one acts against it. Similarly, an embedded attitude should reflect a target's deepest beliefs and is important to them; therefore, acting contrary to that belief is probably surprising.

Second, I propose that targets who act against their committed, embedded views (or moral views, since these may also be committed and embedded) may be seen as lacking in authenticity. *Authenticity* encompasses the even-handed recognition of one's good and bad traits, self-understanding, taking actions based on self-knowledge and striving for sincere and truthful close relationships (Kernis & Goldman, 2006). Someone whose actions contradict their committed, embedded attitudes may lack some of these qualities. Perceivers may feel that such targets lack self-knowledge and fail to act based on self-knowledge; after all, such targets express unwavering, deeply integrated attitudes that they then contradicted. People generally dislike various forms of inauthenticity and feel worse when feeling inauthentic (Crant, 1996; Kifer et al., 2013; Liu & Perrewe, 2006; Markowitz et al., 2023; Portal et al., 2019; Sezer et al., 2018). However, perceivers obviously lack a direct path to knowing a target's authenticity (Bailey & Levy, 2022) and so may rely on indirect cues for appraising targets' authenticity. For example, consistency is crucial for judging authenticity (Eastman, 1994; Kraus et al., 2011; Kreps et al., 2017; Sheldon et al., 1997). Most people endorse a 'strict view' of authenticity in which authentic people's actions should remain consistent with their core values in all situations (Jongman-Sereno & Leary, 2020). Targets' attitude-behaviour inconsistencies, then, should therefore be

² Others have proposed alternative dimensions (e.g., Abelson, 1988; Holland et al., 2003; Krosnick et al., 1993; Prislun, 1996). As Bassili (2008) pointed out, however, most attitude strength dimensions include some form of commitment and certainty regarding attitude position; and some form of embeddedness, ego preoccupation or centrality.

viewed as inauthentic, especially when embedded in the target's core values.

1.2 | Morality as a 'special' attitude strength dimension

Although some work shows moral attitudes to be committed and/or embedded, others argue that moral attitudes have unique properties beyond attitude strength (Mueller & Skitka, 2018; Skitka, 2010; Skitka et al., 2005, 2021). These unique moral attitude qualities may carry important implications for hypocrisy, but only if people draw these distinctions when perceiving others' attitudes (vs. only doing so when conceptualizing their own attitudes). A target who expresses a moral attitude has gone beyond attitude commitment and embeddedness because they imply that the target believes that their views (1) are felt to be more *factual* than non-moral attitudes, understood to be claims about the objective universe; (2) often prompt *intense emotions*, such as anger or disgust at others' violations of one's moral attitudes; (3) stimulate a desire for *social distance* from contrary attitudeholders; and (4) are *universal*, such that everybody should share one's moral attitudes (Skitka et al., 2021). In short, moral attitudes imply that counter-attitudinal action is not just surprising but is a *moral violation*. Indeed, morally convicted people believe that they have no choice and are obliged to follow moral attitudes (Kouchaki et al., 2018; Sabucedo et al., 2018). Insofar as perceivers understand this, they may recognize that moral hypocrites have committed a moral violation by their own standards.

Laurent and Clark (2019) examined how perceivers viewed targets as hypocrites under different conditions of attitude strength. The conditions involved targets having moral views that they contradicted, with variations in whether they imposed their views on others (e.g., 'tells someone emphatically that they shouldn't do X, because it is very wrong'). Perceivers' judgements of hypocrisy were highest when targets acted inconsistently with moral views they imposed on others. This suggests that the universal aspect of moral attitudes strongly influences condemnation of targets who act inconsistently with moral attitudes.

1.3 | The present research

In these experiments, I address multiple questions from previous research. I investigate whether perceivers perceive targets as hypocritical when they act against non-moral attitudes (e.g., eating candy). Additionally, I explore if labelling an attitude as moral before acting against it enhances these judgements. I also examine whether hypocrisy concerning a target's committed or embedded attitudes replicates the effects of moralizing the actor's attitude. Experiments 2 and 3 consider various mechanisms, such as surprise and inauthenticity, to explain why hypocrites against strong attitudes are viewed as particularly negative. Furthermore, Experiment 3 examines whether

moral attitude hypocrites possess distinct characteristics in person perception due to the unique features of moral attitudes compared to other strong attitudes.

1.3.1 | Pilot study

I performed an initial pilot test to determine which topics might be viewed as morally vacuous by most people (hence, most perceivers, who were drawn from the same population in subsequent experiments). Thirty UK Prolific participants were asked about four morally relevant behaviours (e.g., downloading music illegally from the internet) from Jordan et al. (2017), and eight behaviours that I reasoned would be morally vacuous (e.g., eating candies), rating each from 1 (*not at all a moral issue*) to 9 (*entirely a moral issue*). A paired-samples *t*-test confirmed that the non-moral issues ($M = 1.48$, $SD = 0.91$) were seen as irrelevant to morality, whereas the Jordan et al. issues ($M = 5.59$, $SD = 1.26$) were judged to be moderately moral; this difference was substantial, $t(29) = 14.61$, $p < .001$, $d = 2.67$ [1.89, 3.43].

2 | EXPERIMENT 1

In Experiment 1, beyond conceptually replicating past results (Jordan et al., 2017; Teeny et al., 2023), I examined for the first time whether perceivers judge targets more severely when targets contradict targets' own embedded and committed (vs. ordinary) attitudes and compared this to judgements of inconsistencies from moral attitudes. I used morally vacuous topics (e.g., eating candy) that pilot testing had shown were not viewed as inherently moral.

2.1 | Methods

Complete verbatim materials (also see SOM 1–3 in the Supporting Information) and open data/syntax for experiments are openly available at <https://osf.io/whu9z/>. The experiments are not preregistered. All relevant studies, manipulations and measures are reported; no participants' data were discarded.

2.1.1 | Participants

I used a within-participants design to maximize statistical power. I aimed for 100 participants but also retained four additional participants who answered a subset of questions. Based on Prolific's demographic information, participants were 38% males, 61% females, 1% preferred not to answer; 82.5% White, 8.2% mixed, 7.2% Black, 2.1% Asian; and of diverse ages: $\text{range}_{\text{age}} = 19\text{--}71$, $M_{\text{age}} = 39.0$, $SD_{\text{age}} = 13.7$. According to G*Power v. 3.1 (Faul et al., 2009), 100 participants provide 80% power to detect effects of $r > .11$, that is, $d > 0.23$, in my design.

2.1.2 | Procedure and materials

Participants judged seven different vignettes, each between a target person and an interlocuter. Topics were seven non-moral objects from pilot testing: (1) eating candy, (2) taking one's dog to the pub, (3) not downloading the most recent update for one's computer, (4) ordering delivery food, (5) watching TV crime dramas, (6) listening to pop music and (7) drinking sugary beverages. Each vignette was presented in one of several inconsistency conditions.

The first conditions were adapted from Jordan et al. (2017). In the *control* condition, the target simply acted in a particular way without expressing their attitude (e.g., they simply ate some candy). This provided a baseline for how people would appraise someone simply for engaging in the relevant behaviour. In the *behaviourinconsistency* condition (like Jordan et al.'s 'liar' condition), targets stated that they do not perform the behaviour, but then they performed that behaviour (e.g., expressed that they do not eat candies, before eating candies). In the *moralinconsistency* condition, the target stated that they think it is morally wrong to do a behaviour and then they do that behaviour (e.g., saying it morally wrong to eat candies, before eating candies).

In the *attitudeinconsistency* condition, targets stated that they do not like to do a behaviour, but then performed that behaviour (e.g., express that they do not like to eat candies, before eating candies). Participants did not explicitly state that they held a weak (e.g., uncertain) opinion but rather strength was left unspecified (i.e., perceivers would presumably reason that the target had an 'average-strength' opinion given this attitude object).

The remaining three experimental conditions represented my novel extrapolation of Pomerantz et al. (1995) to hypocrisy research: the committed, embedded and combined conditions. The *committedinconsistency* vignettes had the target state that their attitude was 'extreme and 'certain' and that the opinion was 'unlikely to change'. The *embeddedinconsistency* vignettes had the target state that their opinion was 'important to [them] personally' and 'central to [their] identity'. I drew these conditions' wordings directly from Pomerantz et al. (1995). The *combinedinconsistency* condition incorporated elements of both commitment and embeddedness.

To avoid confounding attitude object (e.g., eating candy vs. watching TV crime dramas) with inconsistency type (e.g., control vs. attitude-inconsistent), I created seven between-participant set conditions so that seven of my piloted attitude objects were each paired equally often with each inconsistency type.³ Past hypocrisy studies have sometimes employed similar procedures (e.g., Laurent & Clark, 2019). Participants rated each target immediately after reading its vignette, evaluating how hypocritical they thought the target was, and four evaluation items (i.e., how trustworthy, likeable, honest and good of a person the target was). All items were rated on sliders ranging from 0 (*not at all*) to 100 (*very*) and were derived from Jordan et al. (2017). The evaluation items were internally consistent ($\alpha_s = .90-$

.93 per vignette) and averaged such that high scores indicate a more favourable evaluation of the target.

2.2 | Results and discussion

2.2.1 | Judgements

In this and the remaining experiments, I report hypocrisy and evaluation results together because (low) evaluation scores overwhelmingly mirror (high) hypocrisy scores. Within-subject analysis of variance (ANOVA) revealed that hypocrisy judgements and evaluations each differed based on inconsistency condition. For each dependent variable, I used paired-samples t-tests to follow up with specific planned comparisons, tracked in Table 1. In Figure 1a (hypocrisy) and Figure 1b (evaluations), I display the means for each within-participant condition, and variability is apparent even amongst only the various inconsistency conditions (values to the right of the control condition). Due to an anonymous reviewer's request, contrasts involving the behaviour-inconsistency condition are reported in the Supporting Information (SOM-6) for brevity's sake.

Conceptually replicating past research, moralinconsistency targets were judged as more hypocritical ($M_{\text{hypocrisy}} = 71.90, SD = 33.69$) and less likeable ($M_{\text{likeability}} = 26.73, SD = 16.92$) than control targets ($M_{\text{hypocrisy}} = 27.00, SD = 27.49; M_{\text{likeability}} = 57.92, SD = 16.86$). Thus, neither shifting to a within-participant design nor introducing morally vacuous attitude objects, fundamentally altered people's reactions to these targets (compared to Jordan et al., 2017). This enhances the generalizability of this effect but also reduces concern that these design elements compromised my ability to recover established hypocrisy effects.

Second, I confirmed that attitude-inconsistent targets were seen as more hypocritical ($M_{\text{hypocrisy}} = 59.25, SD = 27.22$) and less likeable ($M_{\text{likeability}} = 37.05, SD = 17.10$) than control participants. These large judgements of hypocrisy and untrustworthiness from a target's acting against a non-moral, non-committed, non-embedded opinion are perhaps surprising given the inanity of the topics (e.g., eating candy, watching TV).

Third, I confirmed that moralinconsistency targets were judged as more hypocritical and were liked less than attitudeinconsistency targets, conceptually replicating Teeny et al. (2023).

Fourth, I began testing novel hypotheses. I examined if committed-inconsistent or embedded-inconsistent targets would be seen as more hypocritical than attitude-inconsistent targets. I expected that these 'strong attitudeinconsistency' conditions would escalate hypocrisy judgements. Committed-inconsistent were seen as slightly more hypocritical ($M_{\text{hypocrisy}} = 69.84, SD = 32.17$) and less likeable ($M_{\text{likeability}} = 27.46, SD = 17.97$) than merely attitude-inconsistent targets. Because the commitment materials allude to the targets being 'certain' (as well as extreme, unlikely to change), these effects generalize Teeny et al.'s (2023) findings that inconsistency from highly certain attitudes is seen as more hypocritical than inconsistency from less certain attitudes. However, note that our comparison point is attitudes of unspecified

³ I arbitrarily discarded another non-moral object, 'leaving the house without an umbrella', because the design only required seven objects to decouple object from condition.

TABLE 1 Effects of inconsistency type on evaluations of targets: Omnibus and planned comparisons. (Experiment 1).

Omnibus test	Hypocrisy			Evaluation		
	$F(6, 582) = 35.27, p < .001, \eta^2 = .27, CI_{90\%} = [0.21, 0.31]$			$F(6, 576) = 90.33, p < .001, \eta_p^2 = .49, CI_{90\%} = [0.43, 0.52]$		
Planned contrast	t-test	p-value	d (CI _{95%})	t-test	p-value	d (CI _{95%})
Moral vs. control	$t(99) = 9.57$	<.001	0.96 [0.72, 1.19]	$t(97) = -13.17$	<.001	-1.33 [-1.60, -1.06]
Attitude vs. control	$t(98) = 8.32$	<.001	0.84 [0.61, 1.06]	$t(99) = -9.49$	<.001	-0.95 [-1.18, -0.71]
Moral vs. attitude	$t(99) = 3.22$.002	0.32 [0.12, 0.52]	$t(98) = -6.51$	<.001	-0.65 [-0.87, -0.44]
Committed vs. attitude	$t(99) = 2.95$.004	0.30 [0.09, 0.49]	$t(100) = -6.05$	<.001	-0.60 [-0.81, -0.39]
Embedded vs. attitude	$t(99) = 1.72$.088	0.17 [-0.03, 0.37]	$t(100) = -5.88$	<.001	-0.59 [-0.80, -0.37]
Combined vs. committed	$t(98) = 1.08$.284	0.11 [-0.09, 0.31]	$t(99) = -3.14$.002	-0.31 [-0.51, -0.11]
Combined vs. embedded	$t(98) = 2.47$.015	0.25 [0.05, 0.45]	$t(100) = -2.49$.015	-0.25 [-0.45, -0.05]
Moral vs. committed	$t(100) = 0.67$.506	0.07 [-0.26, 0.13]	$t(98) = -0.67$.502	-.07 [-0.13, 0.26]
Moral vs. embedded	$t(99) = 1.99$.049	0.20 [0.00, 0.40]	$t(98) = -0.23$.816	-.02 [-0.17, 0.22]
Moral vs. combined	$t(98) = -0.28$.783	-0.03 [-0.17, 0.23]	$t(97) = 2.25$.027	.23 [0.03, 0.43]

certainty rather than Teeny et al.'s (2023) specifying low-certainty opinions as the comparison point.

Intriguingly, however, I also found that embedded-inconsistent targets were judged to be marginally more hypocritical ($M_{\text{hypocrisy}} = 65.07, SD = 36.06$) and liked less ($M_{\text{likeability}} = 27.17, SD = 16.64$) than attitude-inconsistent targets. This demonstrates for the first time that acting against attitudes that one has claimed as identity-relevant, even if non-moral, still escalates the magnitude of moral condemnation that one faces.

Fifth, I tested if combining both features of attitude strength would prompt even greater hypocrisy judgements compared to either attitude strength feature alone. The combined-inconsistent targets ($M_{\text{hypocrisy}} = 73.37, SD = 33.43$) were indeed seen as more hypocritical than the embedded-inconsistent targets, but not more hypocritical than the committed-inconsistent targets. Furthermore, the combined-inconsistent ($M_{\text{likeability}} = 24.38, SD = 17.37$) targets were liked less than both the embedded-inconsistent and the committed-inconsistent targets.

Finally, I tested if moralinconsistency targets were seen as comparably hypocritical to the various 'strong' attitudeinconsistency targets (committed, embedded and especially the combined conditions). moralinconsistency targets were indeed judged as only slightly more hypocritical than embedded-inconsistent targets, and the difference was reduced to approximately zero when moralinconsistency targets were compared against committed-inconsistent or combined-inconsistent.⁴ Furthermore, judges did not like targets who contradicted their committed attitudes any more than those who contra-

dicted their moral convictions, and the same was true for embedded versus moral attitudes.⁵ Combinedinconsistency targets were actually judged more negatively than were moralinconsistency targets, but this effect was not replicated in the methodologically superior Experiment 2 (and per footnote 4-5's indeterminate conclusions from a Bayesian perspective), so I do not consider it further.

In summary, although perceivers saw targets who contradicted the targets' own moral convictions as hypocritical and unlikeable, this social judgement consequence of expressing moral opinions was largely recreated when having targets contradict sufficiently defined 'strong' opinions (per classic attitude strength dimensions; Pomerantz et al., 1995).

2.2.2 | Summary

Beyond replicating a range of past results given the generally novel context of non-moral attitude objects and a within-participant design, Experiment 1 established several novel findings. Broadly, I wanted to test if targets with various kinds of strong attitude (commitment to a given position; embeddedness in identity) would be judged just as harshly as moral hypocrites. My findings supported this conclusion, showing that not only were committed and embedded attitudeholders judged as hypocritical and bad when they transgressed against their opinions, but these effects were very comparable to judgements made against targets who violate their expressed moral convictions. I also found evidence that combining the attitude strength dimensions sometimes produced even stronger judgement effects, particularly on overall likeability. This supports that two classic pillars of attitude strength, commitment and embeddedness, may also be recognized by social perceivers when judging others. This is consistent with the

⁴ Obviously, frequentist tests cannot clearly support the null hypothesis. I, therefore, employed Bayesian one-way repeated measures ANOVA testing (Rouder's method) to estimate the Bayes factor associated with these differences. For context, a comparison of the moralinconsistency versus control conditions overwhelmingly supported that people judged the former more hypocritical, $BF_{10} = 3.29 \times 10^{18}$. In contrast, I found evidence for the null hypothesis when comparing moralinconsistency versus committedinconsistency, $BF_{01} = 10.10$, and when comparing moralinconsistency versus combinedinconsistency, $BF_{01} = 12.20$ (strong evidence according to Lee & Wagenmakers, 2014; also see Jeffreys, 1961). Support for the null was indeterminate when comparing moralinconsistency against embeddedinconsistency, however, $BF_{01} = 1.85$.

⁵ I again found evidence for the null hypothesis when comparing moralinconsistency versus committedinconsistency, $BF_{01} = 10.00$, and when comparing moralinconsistency versus embeddedinconsistency, $BF_{01} = 12.20$, on evaluations. Support for the null was indeterminate when comparing moralinconsistency against combinedinconsistency, however, $BF_{01} = 1.13$.



FIGURE 1 (a) Hypocrisy judgements of targets based on the type of inconsistency. (b) Positive evaluations of targets are based on the type of inconsistency. Error bars capture standard error estimates.

idea that they represent distinct sources of information as their combined influence resulted in harsher judgements than either information provided alone.

3 | EXPERIMENT 2

A limitation of Experiment 1 is that although the attitude objects themselves were pilot tested to be low in moral relevance, targets with

variously 'strong' opinions might have been assumed to have moral attitudes anyway.⁶ For example, attitudes described as central to one's identity in the embedded inconsistency and combined inconsistency conditions might have led participants to assume that this implied higher moral relevance to the target. Therefore, in Experiment 2, I

⁶ Poignantly, Teeny et al. (2023, Study 2) found no evidence that high-certainty (vs. low-certainty) attitudes were assumed to be more morally based, but the absence of evidence is not evidence of absence. Furthermore, perceivers might still assume that embedded attitudes are more morally based (than non-embedded attitudes).

changed all strong attitude conditions (committed, embedded and combined) so that the targets explicitly rejected any moral feeling about the topics.

I also examined several mediators that might help to understand why inconsistency from attitudes is seen as more hypocritical when those attitudes are moral, or non-moral but strong. Teeny et al. (2023) established surprise as one mediator between various strong attitudes (moral and high certainty) and higher hypocrisy ratings. That mediator might also partially account for why embedded attitudes elicit higher hypocrisy ratings because perceivers may not expect targets to violate attitudes they have claimed as part of their identity. I also reasoned that acting inconsistently from identity-relevant attitudes might be viewed as highly inauthentic by perceivers; that is, such inconsistency may suggest a failure to know oneself adequately, which perceivers usually dislike.

3.1 | Methods

3.1.1 | Participants

I attempted to recruit 250 participants and obtained 260 participants with at least partial data, who based on Prolific's demographic information were 44.8% males, 54.4% females, 0.4% prefer not to answer; 90.2% White, 4.1% Asian; 2.4% mixed, 2.0% Black; and of diverse ages: $\text{range}_{\text{age}} = 20\text{--}80$, $M_{\text{age}} = 41.7$, $SD_{\text{age}} = 13.8$. I performed a power analysis using G*Power revealing that 250 participants provided 90% power to detect effects of $d_z > 0.20$, comparing favourably with most effect sizes in Table 1.

3.1.2 | Procedure and materials

The procedure was very similar to Experiment 1. However, the committed, embedded and combined conditions all had the target explicitly state that their opinion was non-moral. For example, the embedded condition had the target state that, 'although it is not a moral issue to [target's name], this opinion is important to [target's name] personally, and central to [his/her] sense of identity'.

Surprise measure

I used two items to capture surprise: 'her/his behaviour was surprising to me', and 'her/his behaviour was unexpected given what she/he had said'. A factor analysis combining these and the inauthenticity items supported that the surprise items loaded on their own dimension. I averaged them so that higher scores indicated more surprise ($\alpha_s = .89\text{--}.96$).

Inauthenticity measure

I developed four items based on Kernis and Goldman's (2006) theory of authenticity. Two items captured a lack of self-knowledge: 'she/he seemed to lack self-awareness', and 'she/he did not seem to know who she/he was'. Two items captured a biased self-view, for example,

'she/he seemed to have a biased view of herself', and 'she/he seemed unable to face her/his own faults'. Based on the prior factor analysis, I averaged these four items into an index where high scores indicated higher perceptions of inauthenticity ($\alpha_s = .89\text{--}.97$).

3.2 | Results and discussion

3.2.1 | Judgements

I first hypothesized and found, via within-subject ANOVA, that both hypocrisy judgements and evaluations differed based on inconsistency condition. I track specific comparisons in Table 2. In Figure 2a (hypocrisy) and Figure 2b (evaluations), I display the means for each within-participant condition. Again, moralinconsistency targets were judged as more hypocritical ($M_{\text{hypocrisy}} = 79.03$, $SD = 30.31$) and less likeable ($M_{\text{likeability}} = 25.47$, $SD = 18.82$) than were control targets ($M_{\text{hypocrisy}} = 11.66$, $SD = 20.78$; $M_{\text{likeability}} = 60.13$, $SD = 20.38$). Attitude-inconsistent targets were also deemed more hypocritical ($M_{\text{hypocrisy}} = 62.70$, $SD = 30.15$) and less likeable ($M_{\text{likeability}} = 37.83$, $SD = 17.89$) relative to control targets. Moralinconsistency targets were judged to be more hypocritical and were more disliked than attitudeinconsistency targets. This replicates Jordan et al. (2017) and my Experiment 1.

I hypothesized and confirmed that committed-inconsistent ($M_{\text{hypocrisy}} = 72.37$, $SD = 30.02$) and embedded-inconsistent ($M_{\text{hypocrisy}} = 72.91$, $SD = 29.00$) targets would be seen as more hypocritical than attitude-inconsistent targets. Furthermore, both committed-inconsistent ($M_{\text{likeability}} = 31.20$, $SD = 18.73$) and embedded-inconsistent targets ($M_{\text{likeability}} = 30.61$, $SD = 18.97$) were disliked more than were merely attitude-inconsistent targets. Critically, this supports that Experiment 1's effects were not reliant on people mistaking commitment or embeddedness as automatically implying morality, since the present effects were if anything stronger than in Experiment 1, despite explicitly stating that embedded and committed targets' views were *not* morally based. This again demonstrates that beyond features implying that the target's views are committed to a particular position, perceivers respond to how embedded the target's views are in their identity.

I tested if combining both features of attitude strength would prompt even greater hypocrisy judgements compared to either attitude strength feature alone. The combined-inconsistent target was only viewed as no more hypocritical ($M_{\text{hypocrisy}} = 75.88$, $SD = 31.09$) than the embedded-inconsistent targets, and only marginally more hypocritical than the committed-inconsistent targets. However, the combined-inconsistent was liked less ($M_{\text{likeability}} = 26.62$, $SD = 18.04$) than both the embedded-inconsistent and the committed-inconsistent targets.

Finally, I tested if moralinconsistency targets were seen as differently hypocritical than the various 'strong' (committed, embedded and combined) attitude inconsistency targets. Moralinconsistency targets were judged as only slightly more hypocritical than commitment-inconsistent and embedded-inconsistent targets and only marginally

TABLE 2 Effects of inconsistency type on evaluations of targets: Omnibus and planned comparisons (Experiment 2).

Omnibus test	Hypocrisy			Evaluation		
	$F(6, 1380) = 201.08, p < .001, \eta^2 = .47, CI_{90\%} = [0.44, 0.49]$			$F(6, 1470) = 187.23, p < .001, \eta^2 = .43, CI_{90\%} = [0.40, 0.46]$		
Planned contrast	t-test	p-value	d (CI _{95%})	t-test	p-value	d (CI _{95%})
Moral vs. control	$t(233) = 26.72$	<.001	1.75 [1.54, 1.95]	$t(247) = -22.03$	<.001	-1.40 [-1.57, -1.22]
Attitude vs. control	$t(233) = 21.13$	<.001	1.38 [1.20, 1.56]	$t(250) = -15.25$	<.001	-.96 [-1.11, -0.81]
Moral vs. attitude	$t(251) = 8.37$	<.001	0.53 [0.40, 0.66]	$t(248) = -11.97$	<.001	-.76 [-0.90, -0.62]
Committed vs. attitude	$t(250) = 4.70$	<.001	0.30 [0.17, 0.42]	$t(247) = -5.74$	<.001	-0.36 [-0.49, -0.24]
Embedded vs. attitude	$t(251) = 5.24$	<.001	0.33 [0.20, 0.46]	$t(249) = -6.56$	<.001	-0.42 [-0.54, -0.29]
Combined vs. committed	$t(251) = 1.69$.092	0.11 [-0.02, 0.23]	$t(248) = -4.43$	<.001	-0.28 [-0.41, -0.15]
Combined vs. embedded	$t(253) = 1.64$.102	0.10 [-0.02, 0.23]	$t(250) = -4.29$	<.001	-0.27 [-0.40, -0.14]
Moral vs. committed	$t(251) = 3.56$	<.001	0.22 [0.10, 0.35]	$t(246) = -5.73$	<.001	-0.37 [-0.49, -0.24]
Moral vs. embedded	$t(252) = 3.77$	<.001	0.24 [0.11, 0.36]	$t(248) = -4.89$	<.001	-0.31 [-0.44, -0.18]
Moral vs. combined	$t(252) = 1.91$.057	0.12 [-0.004, 0.24]	$t(248) = -1.13$.258	-0.07 [-0.20, 0.05]

more hypocritical than combined-inconsistent.⁷ Turning to evaluations, I found that moral inconsistency targets were liked less than committed inconsistency and than embedded inconsistency targets. However, moral inconsistency targets were not evaluated any more negatively than were combined inconsistency targets.⁸ Thus, despite explicitly stating that the target's committed and embedded views were not moral, people viewed such a target as comparably hypocritical and unlikeable to a moral hypocrite.

3.2.2 | Mediators

I next considered the possible mediators: surprise and inauthenticity (see Table 3).⁹ I first hypothesized and found, via within-subject ANOVA, that surprise differed based on inconsistency condition. Similarly, judgements of the target's inauthenticity differed by condition. For brevity's sake, and because the patterns of means for these variables were generally comparable to the hypocrisy and evaluation variables, I summarize these results briefly and plot them in Figure 3a,b. In short, targets were judged as more surprising and more inauthentic, given (1) moral inconsistency > control, (2) attitude inconsistency > control, (3) moral inconsistency > attitude inconsistency, (4) committed- or embedded inconsistency > attitude inconsistency, (5)

combined inconsistency > embedded- or committed inconsistency and (6) moral inconsistency > embedded- or committed inconsistency (all $ps < .009$). Importantly, however, moral inconsistency targets were neither judged as more surprising, $t(252) = -0.01, p = .832, d = -0.01 [-0.14, 0.11]$, nor more inauthentic, $t(252) = -0.42, p = .674, d = 0.03 [-0.10, 0.15]$, than combined inconsistency targets.¹⁰

Next, using the lavaan package (Rosseel, 2012) in R (R Core Team, 2022), I tested if the committed, embedded and/or moral conditions' effects on evaluation were accounted for by elevated surprise and/or inauthenticity judgements. To do this, I constructed contrast codes comparing these conditions against the attitude inconsistency condition. Then I conducted three multilevel mediations, which simultaneously model level 1 mediation (i.e., the extent to which, within participants, the committed or embedded condition's effects on evaluation were accounted for by surprise and/or inauthenticity) and level 2 regression (i.e., the extent to which participants who generally saw targets as more surprising or inauthentic generally liked targets more or less). I focus on the level 1 mediation effects, tracked in Table 4.

Results were very similar when assessing each of the strong attitude conditions. Targets who expressed and contradicted embedded, committed or moral attitudes were judged more surprising and more inauthentic than attitude-inconsistent targets (first data column). Surprise and inauthenticity were related to less favourable judgements of targets in all analyses (second data column). Additionally, the effects of embedded (committed) conditions versus attitude inconsistency on evaluation were reduced to non-significance by the addition of the mediators (direct effects, respectively: $Z = -1.73, p = .083; Z = -1.69, p = .091$). In contrast, the effect of moral inconsistency (vs. attitude inconsistency) on evaluations remained substantial even with the mediators included (direct effect: $Z = -5.16, p < .001$). Finally, both surprise and inauthenticity showed, in parallel, significant evidence of mediating

⁷ Interestingly, unlike Experiment 1, Bayesian analysis revealed 'strong' to 'very strong' support for the alternative hypothesis that the moral inconsistency targets were seen as more hypocritical than the commitment-inconsistency, $BF_{10} = 23.24$, and than the embedded inconsistency, $BF_{10} = 47.42$. However, there was also moderate support for the null hypothesis that the moral inconsistency targets were seen as equally hypocritical as the combined inconsistency targets, $BF_{01} = 3.31$.

⁸ Bayesian testing gave 'extreme' support for the alternative hypothesis that moral inconsistent targets were disliked more than committed inconsistency, $BF_{10} = 224,958.7$ and disliked more than embedded inconsistency targets, $BF_{10} = 4087.5$. However, I also found 'strong' support for the null hypothesis that the moral inconsistency targets were seen as equally unlikeable compared with combined inconsistency targets, $BF_{01} = 10.8$.

⁹ Power analysis for multilevel mediation depends on a broad range of factors for which it is difficult to provide meaningful a priori expectations (e.g., see Zigler & Ye, 2019). Thus, these mediations may be considered exploratory, with Experiment 3 providing a confirmatory replication.

¹⁰ Again, Bayesian testing supported that moral-inconsistent targets were not seen as more surprising, $BF_{01} = 18.2$, or inauthentic, $BF_{01} = 18.9$, than combined-inconsistent targets.

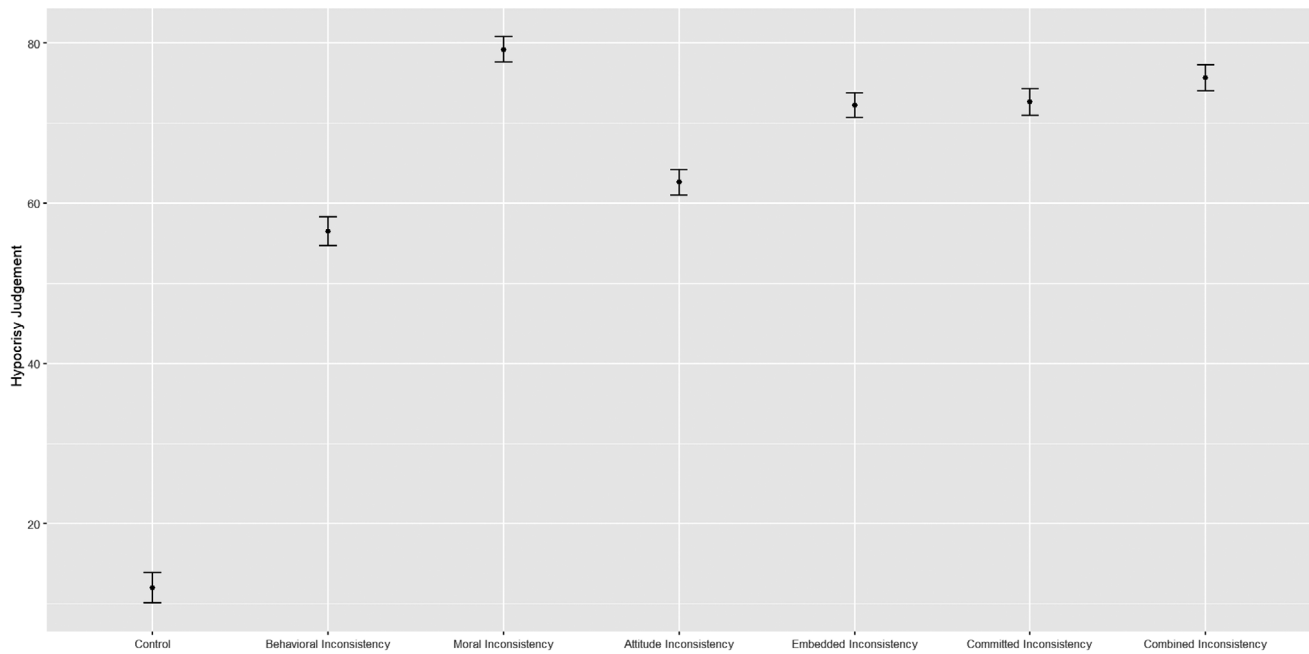
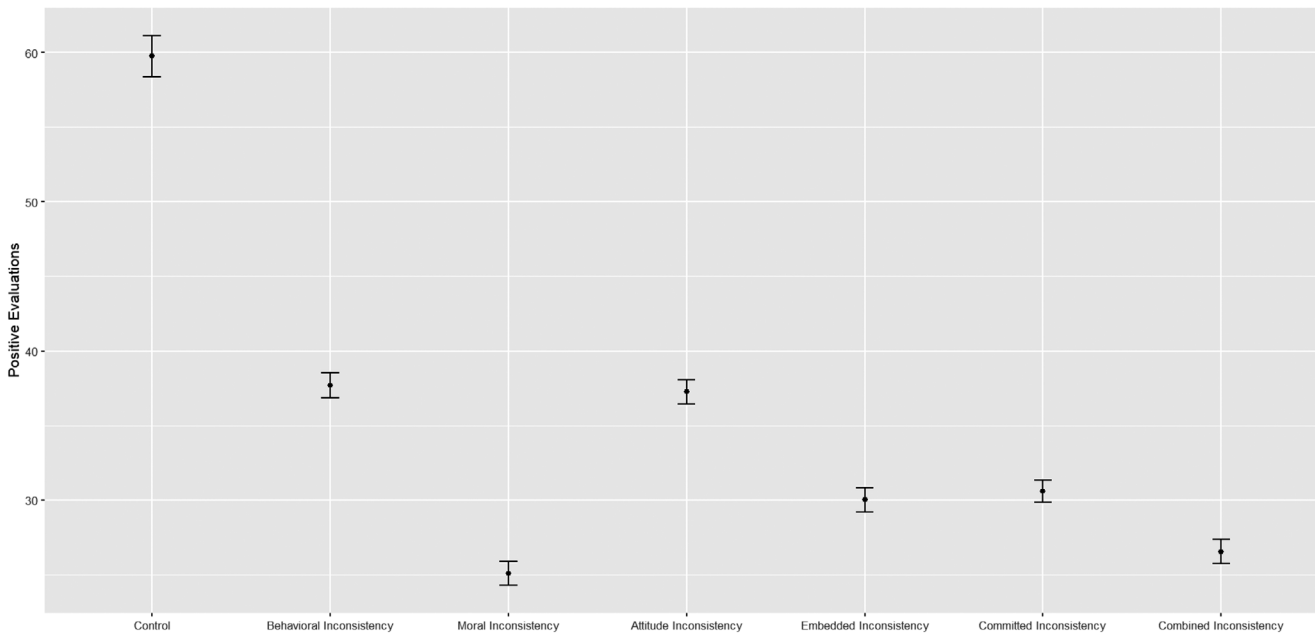
(a) *Hypocrisy Judgements of Targets Based on the Type of Inconsistency (Experiment 2)*(b) *Positive Evaluations of Targets Based on the Type of Inconsistency (Experiment 2)*

FIGURE 2 (a) Hypocrisy judgements of targets based on the type of inconsistency (Experiment 2). (b) Positive evaluations of targets based on the type of inconsistency (Experiment 2). Error bars capture standard error estimates.

the effects of embedded, committed and moral (vs. regular) attitude inconsistency on evaluations.

3.2.3 | Summary

Experiment 2 explored hypocrisy judgements, replicating Experiment 1's findings on target commitment and embeddedness effects on

perceiver's judgements. Moreover, it did so despite all the 'strong' attitude inconsistency conditions (committed, embedded and combined) emphasizing that targets had non-moral opinions, eliminating a counter-explanation of Experiment 1. Nonetheless, targets who contradicted attitudes that were certain and unwavering (committed) or personally invested (embedded) were judged to be just as hypocritical, unlikeable, surprising and inauthentic as those who transgressed against their moral convictions. Again, this occurred despite using

TABLE 3 Effects of inconsistency type on mediators: Omnibus and planned comparisons (Experiment 2).

Omnibus test	Surprise			Inauthenticity		
	$F(6, 1506) = 463.52, p < .001, \eta^2 = .65, CI_{90\%} = [0.63, 0.67]$			$F(6, 1506) = 293.60, p < .001, \eta^2 = .54, CI_{90\%} = [0.51, 0.56]$		
Planned contrast	t-test	p-value	d (CI _{95%})	t-test	p-value	d (CI _{95%})
Moral vs. control	$t(251) = 32.06$	<.001	2.02 [1.80, 2.23]	$t(251) = 26.48$	<.001	1.67 [1.48, 1.86]
Attitude vs. control	$t(251) = 26.91$	<.001	1.70 [1.50, 1.89]	$t(251) = 20.46$	<.001	1.29 [1.12, 1.46]
Moral vs. attitude	$t(252) = 10.05$	<.001	0.63 [0.50, 0.77]	$t(252) = 10.38$	<.001	0.65 [0.52, 0.79]
Committed vs. attitude	$t(252) = 6.60$	<.001	0.42 [0.29, 0.54]	$t(252) = 6.72$	<.001	0.42 [0.29, 0.55]
Embedded vs. attitude	$t(252) = 7.59$	<.001	0.48 [0.35, 0.61]	$t(252) = 8.06$	<.001	0.51 [0.38, 0.64]
Combined vs. committed	$t(252) = 2.68$.008	0.17 [0.04, 0.29]	$t(252) = 4.06$	<.001	0.26 [0.13, 0.38]
Combined vs. embedded	$t(252) = 3.24$.001	0.20 [0.08, 0.33]	$t(252) = 3.90$	<.001	0.25 [0.12, 0.37]
Moral vs. committed	$t(252) = 3.31$.001	0.21 [0.08, 0.33]	$t(252) = 4.15$	<.001	0.26 [0.14, 0.39]
Moral vs. embedded	$t(252) = 3.27$.001	0.21 [0.08, 0.33]	$t(252) = 3.51$.001	0.22 [0.10, 0.39]
Moral vs. combined	$t(252) = 0.42$.674	0.03 [-0.10, 0.15]	$t(252) = -0.21$.832	-0.01 [-0.14, 0.11]

morally vacuous attitude objects (e.g., eating candy, ordering delivery food).

Furthermore, I found that inconsistency in behaviour not only leads to dislike due to surprise (previously established mediator; Teeny et al., 2023) but also because it is perceived as lacking authenticity. Specifically, perceivers judged that such targets may lack self-awareness (i.e., not know who they are) and may have a biased self-perception (i.e., acknowledge only their favourable characteristics), making them unlikeable. Interestingly, both mechanisms mediated the effects for committed, embedded and moral attitudes. This integrates hypocrisy research with a broader literature on perceived authenticity (Bailey & Levy, 2022; Kernis & Goldman, 2006; Newman, 2019; Sutton, 2020) in which inaccurate self-knowledge prompts disliking (Garcia, 2019; Liu & Perrewe, 2006).

Finally, although inconsistency from committed and embedded attitudes affected evaluations through similar mechanisms as inconsistency from moral attitudes, the latter showed a much larger direct effect on evaluations after adjusting for the mediators. Moral hypocrisy might prompt dislike for reasons beyond surprise and inauthenticity, whereas hypocrisy from merely 'strong' attitudes produces dislike only for these reasons. Thus, in Experiment 3 I examined whether moral and other kinds of 'strong' attitudes might differ in some respects.

4 | EXPERIMENT 3

In Experiments 1 and 2, I focused on various conditions under which people are labelled as hypocritical and unlikeable for acting in ways that are inconsistent with their attitudes, and some reasons why perceivers form these evaluations. In Experiment 3, I consider under what conditions perceivers may be led to discard hypocrisy labels, such as when they consider more situational explanations for others' behaviour. Beyond their attitudes, people's behaviour is driven by past behaviours/habits, perceived behavioural control, moral norms, self-identity and so on (e.g., Ajzen, 1991; Conner & Armitage, 1998; Kan

& Fabrigar, 2017). For example, someone who dislikes eating candy might, nonetheless, eat some because of social pressure (e.g., everyone else is eating it), because of a lack of perceived behavioural control (e.g., because they have no other food and are hungry) or other transitory influences. However, perceivers often fail to consider how these situational determinants may shape other people's behaviours (Gilbert & Malone, 1995; Heider, 1958; Ross, 1977, 2001), though they are more likely to do when given sufficient cognitive resources (Gilbert et al., 1988).

As Gawronski (2004) articulated in his work on theory-based corrections, people may fail to adequately accommodate situational explanations in their attributions (1) because they lack sufficient knowledge of a situational lay theory, (2) because they fail to adequately apply situational theory to a given target, (3) because they are deliberately neglecting situational theory and (4) because they may apply situational theory in a way that promotes dispositional inferences. People must be motivated and able to apply situational reasoning to correct for an excessively disposition-focused attribution pattern. I think that perceivers may generally be less inclined to attribute attitude-inconsistent behaviour to the target being 'a hypocrite' (i.e., a trait attribution) if they are given sufficient motivation and opportunity to consider situational factors for others' behaviour.

Experiments 1 and 2 mostly show that moral inconsistency is similar to sufficiently strong attitude inconsistency – even morally vacuous issues. Nonetheless, I propose that inconsistency in one's moral convictions will sometimes produce unique effects among social perceivers. Specifically, moral convictions might be less affected by situational explanations relative to other strong-attitude inconsistencies. Recall that unique qualities may be entailed by one's moral (vs. non-moral) beliefs, including that moral convictions are universal, objective and deeply emotional. These morality-specific properties might have implications for hypocrisy judgements.

First, by declaring a moral attitude, the target has judged the perceiver (and everyone else) insofar as the perceiver might act against the attitude object, and yet that target has acted in this contrary manner

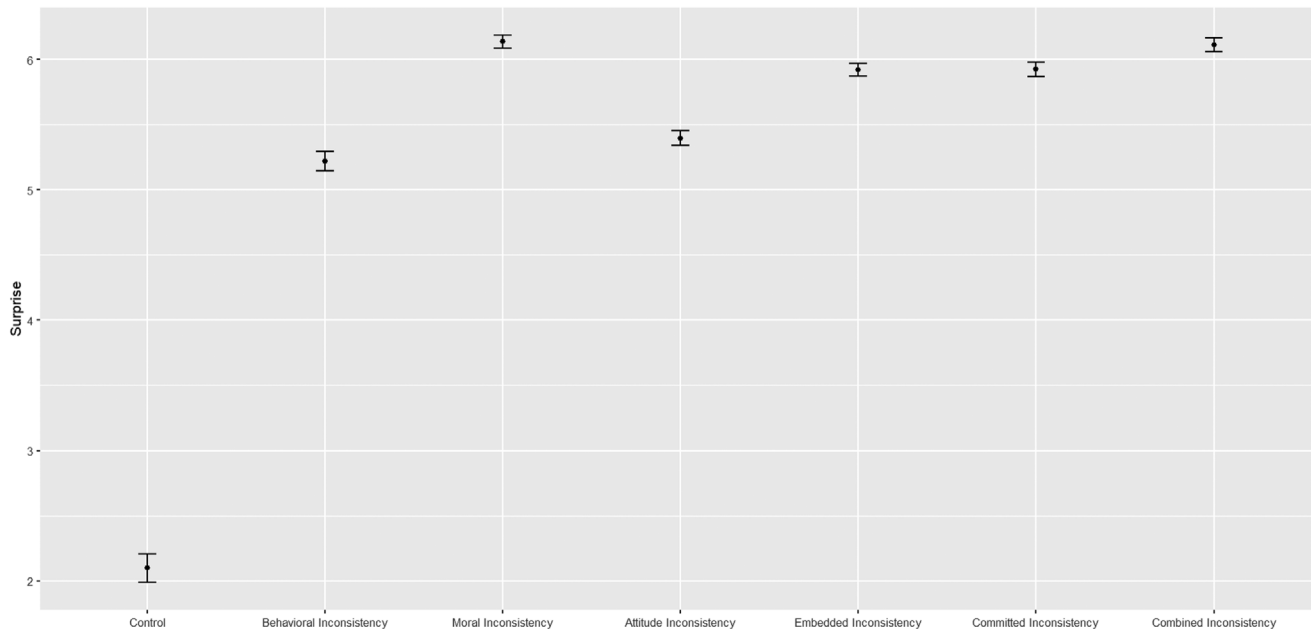
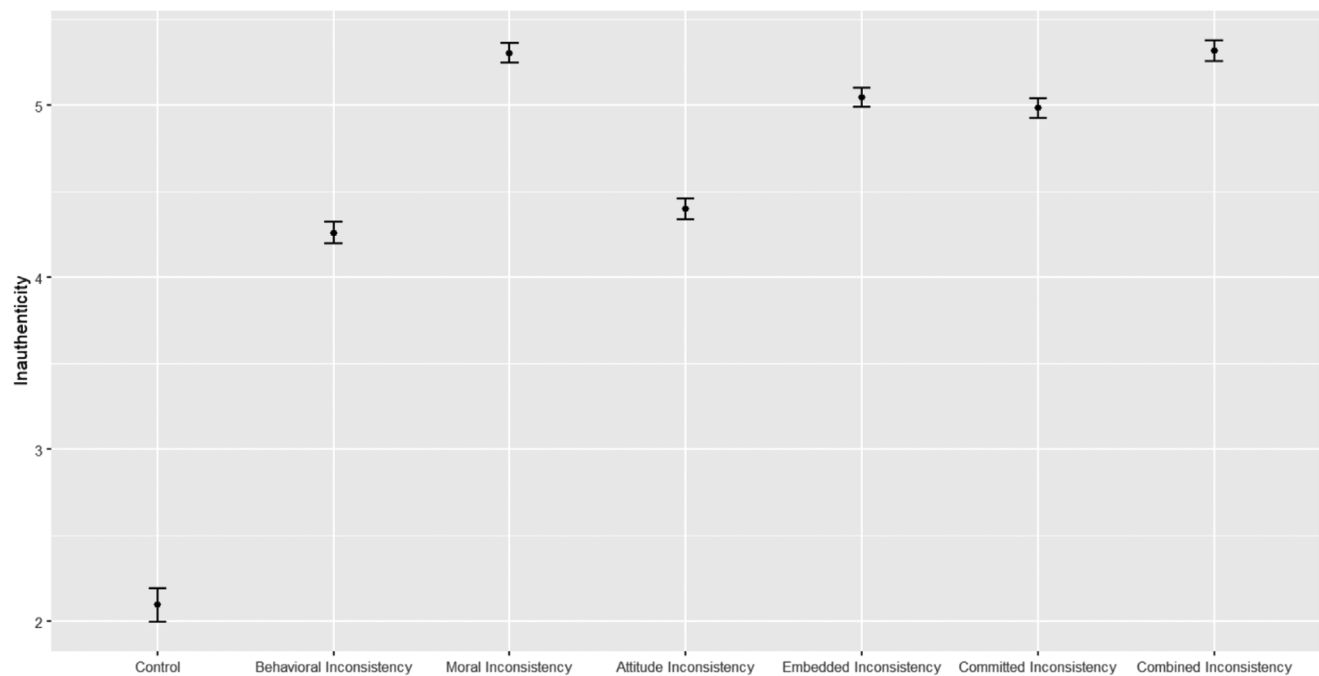
(a) *Surprise in Reaction to Each Type of Inconsistency (Experiment 2)*(b) *Inauthenticity of Target Based on the Type of Inconsistency (Experiment 2)*

FIGURE 3 (a) Surprise in reaction to each type of inconsistency (Experiment 2). (b) Inauthenticity of target based on the type of inconsistency (Experiment 2). Error bars capture standard error estimates.

themselves. For instance, the target claims that ordering delivery food is a moral wrong, then does this themselves. Many perceivers will feel judged by the opening claim insofar as the perceiver has themselves done this action, but not feel morally judged when a target expresses a strong, non-moral opinion. Consequently, perceivers' motivation to apply situational theory to such hypocrites may be sabotaged, producing dispositional judgements (Gawronski, 2004).

Second, because moral attitudes are supposed to be universal, objective claims, a perceiver may believe that there are no situational considerations that undo the hypocrisy of the act. By the target's implying that their attitude position is an objective fact that must be universally upheld, perceivers may feel that situational considerations are not insufficient to 'forgive' the target's behaviour. The target is 'obliged' to act on their moral attitudes (Kouchaki et al.,

TABLE 4 Inconsistency type predicts evaluations via surprise and inauthenticity: Within-participant effects (Experiment 2).

Committed vs. attitude		Mediator to evaluation	Indirect effect	Z-value of indirect effect
Mediator	Condition to mediator			
Surprise	$\beta = .20, SE = 0.04, z = 4.70, p < .001$	$\beta = -.31, SE = 0.07, z = -4.31, p < .001$	$-0.06, SE = 0.02$	$-3.18, p = .001$
Inauthenticity	$\beta = .21, SE = 0.04, z = -4.96, p < .001$	$\beta = -.48, SE = 0.07, z = 7.02, p < .001$	$-0.10, SE = 0.03$	$-4.05, p < .001$
Embedded vs. attitude				
Mediator	Condition to mediator	Mediator to evaluation	Indirect effect	Z-value of indirect effect
Surprise	$\beta = .20, SE = 0.04, z = 4.83, p < .001$	$\beta = -.37, SE = .08, z = -4.72, p < .001$	$-0.08, SE = 0.02$	$-3.38, p < .001$
Inauthenticity	$\beta = .24, SE = 0.04, z = -5.69, p < .001$	$\beta = -.46, SE = .07, z = 6.23, p < .001$	$-0.11, SE = 0.03$	$-4.19, p < .001$
Moral vs. attitude				
Mediator	Condition to mediator	Mediator to evaluation	Indirect effect	Z-value of indirect effect
Surprise	$\beta = .29, SE = 0.04, z = 7.20, p < .001$	$\beta = -.15, SE = 0.07, z = -2.05, p = .041$	$-0.04, SE = 0.02$	$-1.97, p = .049$
Inauthenticity	$\beta = .32, SE = 0.04, z = 8.02, p < .001$	$\beta = -.60, SE = 0.06, z = 9.50, p < .001$	$-0.19, SE = 0.03$	$-6.19, p < .001$

2018; Sabucedo et al., 2018). Again, this would presumably erode perceivers' motivation to depart from dispositional judgements of the moral hypocrite.

4.1 | Methods

I reasoned that replicating and extending my findings with a fully between-participant design experiment would be desirable, so I did this in Experiment 3.

4.1.1 | Participants

I aimed to recruit 400 participants but oversampled to $N = 434$ due to some missing data in 34 participants. Analyses thus have varying degrees of freedom. Based on Prolific's demographic information, participants were 38.5% males, 61.3% females, 0.3% prefer not to answer; 89.2% White, 5.8% Asian, 2.3% Black, 2.0% mixed; and of diverse ages: $range_{age} = 18-79$, $M_{age} = 41.9$, $SD_{age} = 13.9$. Four hundred participants give 80% power to detect effects of $r > .16$ ($d > 0.33$). I examined statistical power for various 'shapes' of the critical interaction, but this process was inconclusive and so my sample size was determined by a '50/cell' goal set by my financial resources (see SOM-5 in the Supporting Information). I was 80% powered to replicate most of Experiment 2's mediation effects (and these were indeed replicated in this confirmatory sample; see SOM-7 in the Supporting Information).

4.1.2 | Procedure and materials

The design was a 4 (inconsistency type: control vs. moralinconsistency vs. attitudeinconsistency vs. combinedinconsistency) \times 2 (attribution style: dispositional vs. situational) between-participants design. All participants read a vignette about two friends discussing a topic (eating candy), and then one interlocuter eats some candy. As a reduced version of the prior experiments, I included four inconsistency conditions: the target (Kevin) either expressed no opinion about the topic (*control*), expressed his moral opposition to this behaviour (*moralinconsistency*), expressed that he did not like to eat candy (*attitudeinconsistency*) or expressed a committed and embedded but explicitly non-moral opinion that he did not like eating candy (*combinedinconsistency*).

Participants were next assigned to one of two conditions. In the *dispositional attribution* condition, participants thought about what unchanging traits of Kevin might have led him to behave as he did. In the *situational attribution* condition, participants thought about what contextual/situational factors might have caused Kevin's behaviour. Either way, participants thought for 60 seconds before the study advanced them to the dependent variables. Inspection of these essays confirmed that those in the dispositional (situational) conditions gave appropriate responses. 'Dispositional' participants said that an inconsistent (any non-control) Kevin was a liar, a hypocrite, desperate for attention, etc. 'Situational' participants said that an inconsistent Kevin might have

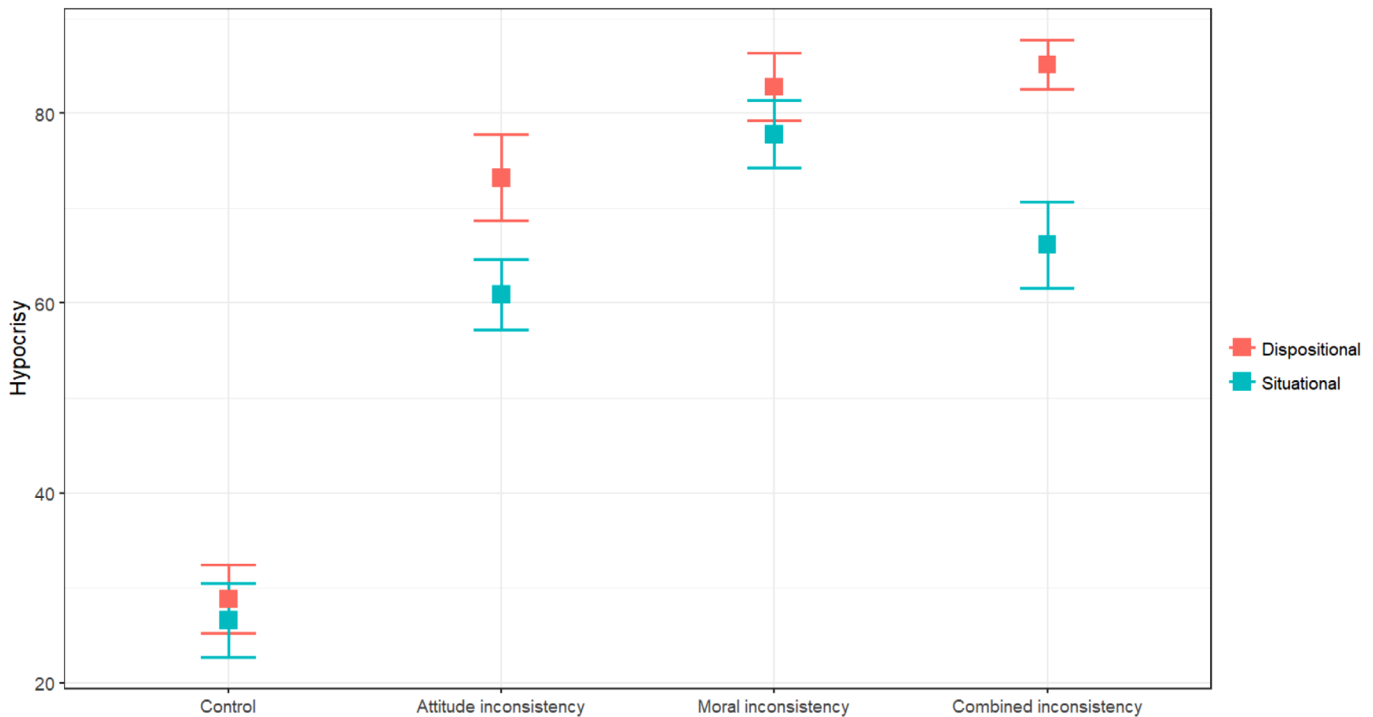


FIGURE 4 Attribution style attenuates the hypocrisy-increasing effects of even strong attitude inconsistency, but not moral inconsistency. Error bars capture standard errors.

only had candy to eat, disliked the third party with whom they had been discussing candy, etc.

I worded measures as per Experiment 2 but added four additional items. One question was intended as a manipulation check: 'Kevin's opinion about candy is based on his moral beliefs'. The remaining three items probed Skitka and colleagues' (2005) ways in which moral attitudes differ from other strong attitudes. The *objectivity* item asked if participants thought 'Kevin thinks of his opinion of candy-eating as factual'. The *universalism* item asked if 'Kevin thinks that everybody should agree with him about eating candy'. The *emotions* item asked if 'Kevin likely experiences intense emotions related to his opinions about eating candy'.

4.2 | Results and discussion

4.2.1 | Hypocrisy judgements

I used a 4 (inconsistency type) \times 2 (attribution style) ANOVA to examine factors contributing to judgements of Kevin's hypocrisy. I detected main effects of inconsistency type, $F(3, 388) = 77.38, p < .001, \eta_p^2 = .37$ and of attribution style, $F(1, 388) = 13.08, p < .001, \eta_p^2 = .03$. The critical interaction, however, should qualify these effects. Although this effect was non-significant, $F(3, 388) = 1.98, p = .117, \eta_p^2 = .02$, I examined if at least the difference between moral inconsistency and combined inconsistency was altered by thinking situationally versus dispositionally. This simpler 2 \times 2 interaction was significant, $F(1, 197) = 3.91, p = .049, \eta_p^2 = .02$.

As captured in Figure 4, if Kevin expressed a committed and embedded (but non-moral) view that he then contradicted with his behaviour (i.e., combined condition), people's views of his hypocrisy dropped when they were prompted to consider situational factors, $t(97) = 3.79, p < .001, d = 0.77 [0.35, 1.18]$. But if Kevin expressed a moral view, the stimulation of situational thinking did not reduce hypocrisy judgements, $t(100) = 1.00, p = .322, d = 0.20 [-0.19, 0.59]$.¹¹ Alternatively, engaging in dispositional thinking led perceivers to similar hypocrisy judgements for the combined inconsistency and moral inconsistency conditions, $t(105) = 0.54, p = .590, d = 0.10 [-0.28, 0.48]$; however, perceivers led to think situationally judged combined inconsistency Kevin as significantly less hypocritical than moral inconsistency Kevin, $t(92) = -2.04, p = .044, d = -0.42 [-0.83, -0.01]$.

4.2.2 | Other evaluations

Turning to evaluation, I found a main effect of inconsistency type. Specific contrasts, this time analysed as between-participant t -tests, are tracked in Table 5. Kevin was liked more in the control ($M_{\text{likeability}} = 53.27, SD = 16.41$), less liked in the attitude inconsistency ($M_{\text{likeability}} = 36.81, SD = 15.99$), and even less so in the moral inconsistency ($M_{\text{likeability}} = 30.49, SD = 18.06$) versus attitude inconsistency condition. The moral inconsistency condition, however, did not differ from the

¹¹ As an anonymous reviewer noted, with a sufficiently large sample size this simple slope could be detectably positive. Nonetheless, the significant interaction indicates that the slope is attenuated relative to the combined condition, which is what my reasoning predicted.

TABLE 5 Effects of inconsistency type on evaluations of targets: omnibus and planned comparisons (Experiment 3).

Hypocrisy			
Omnibus test	$F(3, 388) = 77.38, p < .001, \eta^2 = .37, CI_{90\%} = [0.31, 0.43]$		
Planned contrast	t-test	p-value	d (CI _{95%})
Attitude vs. control	$t(392) = 9.94$	<.001	0.71 [0.56, 0.86]
Moral vs. attitude	$t(392) = 3.80$	<.001	0.27 [0.13, 0.40]
Moral vs. combined	$t(392) = 0.96$.336	0.07 [-0.07, 0.21]
Evaluation			
Omnibus test	$F(3, 390) = 47.10, p < .001, \eta^2 = .27, CI_{90\%} = [0.20, 0.32]$		
Planned contrast	t-test	p-value	d (CI _{95%})
Attitude vs. control	$t(394) = -6.95$	<.001	-0.49 [-0.63, -0.35]
Moral vs. attitude	$t(394) = -2.70$.007	-0.19 [-0.33, -0.05]
Moral vs. combined	$t(394) = 0.86$.391	0.06 [-0.08, 0.20]
Surprise			
Omnibus test	$F(3, 390) = 77.06, p < .001, \eta^2 = .37, CI_{90\%} = [0.31, 0.42]$		
Planned contrast	t-test	p-value	d (CI _{95%})
Attitude vs. control	$t(394) = 10.88$	<.001	0.77 [0.62, 0.92]
Moral vs. attitude	$t(394) = 1.95$.052	0.14 [-0.001, 0.27]
Moral vs. combined	$t(394) = -0.68$.499	-0.05 [-0.19, 0.09]
Inauthenticity			
Omnibus test	$F(3, 390) = 63.20, p < .001, \eta^2 = .33, CI_{90\%} = [0.26, 0.38]$		
Planned contrast	t-test	p-value	d (CI _{95%})
Attitude vs. control	$t(394) = 8.80$	<.001	0.62 [0.48, 0.77]
Moral vs. attitude	$t(394) = 3.23$.001	0.23 [0.09, 0.36]
Moral vs. combined	$t(394) = 0.11$.913	0.01 [-0.13, 0.15]

combined inconsistency ($M_{\text{likeability}} = 28.44, SD = 16.41$) condition.¹² Additionally, I detected a main effect of attribution style, $F(1, 390) = 22.54, p < .001, \eta_p^2 = .06$, indicating that people simply liked Kevin better when they were using situational ($M_{\text{likeability}} = 40.69, SD = 19.00$) rather than dispositional thinking ($M_{\text{likeability}} = 33.75, SD = 18.93$).

For surprise, I found a main effect only of the inconsistency condition whereby Kevin surprised people less in the control ($M_{\text{surprise}} = 2.98, SD = 1.61$) than in the attitude inconsistency ($M_{\text{surprise}} = 5.31, SD = 1.56$). Kevin was also viewed as marginally more surprising in the moral inconsistency ($M_{\text{surprise}} = 5.72, SD = 1.51$) versus the attitude inconsistency condition. The moral inconsistency condition, however, did not differ from the combined inconsistency ($M_{\text{surprise}} = 5.86, SD = 1.33$) condition.¹³ Attribution style had neither main nor interaction effects on surprise, $F_s < 2.61, p_s > .107$.

Finally, for inauthenticity, I found a main effect of Inconsistency type, such that Kevin was seen as more inauthentic in the attitude inconsistency ($M_{\text{inauthenticity}} = 4.49, SD = 1.23$) than in the control

condition ($M_{\text{inauthenticity}} = 2.96, SD = 1.44$). Kevin was also judged as being more inauthentic in the moral inconsistency ($M_{\text{inauthenticity}} = 5.05, SD = 1.08$) versus the attitude inconsistency condition. The moral inconsistency condition, however, did not differ from the combined inconsistency ($M_{\text{inauthenticity}} = 5.03, SD = 1.12$) condition.¹⁴ Additionally, I detected a main effect of attribution style, $F(1, 390) = 12.24, p = .001, \eta_p^2 = .03$, indicating that perceivers thought Kevin more inauthentic when they used dispositional ($M_{\text{inauthenticity}} = 4.57, SD = 1.52$) rather than situational attributions ($M_{\text{inauthenticity}} = 4.21, SD = 1.42$).

4.2.3 | Unique moral qualities

I next considered whether morally inconsistent Kevin was evaluated differently compared to the two attitude inconsistency Kevins, to see if perceivers infer the sorts of differences that have been attributed to moral attitudes (i.e., objectivity, universalism and emotionality; see Table 6). I found marginal evidence that the morally inconsistent Kevin was seen as having a more morally rooted opinion than his normal

¹² A Bayesian t-test using Rouder's method and Jeffrey's default priors revealed 'moderate' evidence for the null hypothesis that the moral inconsistency and combined inconsistency Kevins were liked equally, $BF_{01} = 6.43$.

¹³ I found 'moderate' evidence for the null hypothesis that the moral inconsistency and combined inconsistency Kevins were equally surprising, $BF_{01} = 7.03$.

¹⁴ I found 'moderate' evidence for the null hypothesis that the moral inconsistency and combined inconsistency Kevins were judged to be equally inauthentic, $BF_{01} = 8.95$.

TABLE 6 Effects of inconsistency type on perceived target levels of qualities proposed to reflect moral attitudes (Experiment 3).

	Moral	Objectivity	Universalism	Emotions
Omnibus test	$F(3, 390) = 4.87, p = .002, \eta^2 = .04, CI_{90\%} = [0.01, 0.07]$	$F(3, 390) = 4.43, p = .004, \eta^2 = .03, CI_{90\%} = [0.01, 0.06]$	$F(3, 390) = 21.75, p < .001, \eta^2 = .14, CI_{90\%} = [0.09, 0.19]$	$F(3, 390) = 11.24, p < .001, \eta^2 = .08, CI_{90\%} = [0.04, 0.12]$
Control	2.64 (1.40)	3.68 (1.66)	2.92 (1.68)	3.09 (1.69)
Attitudeinconsistency	2.98 (1.74)	4.01 (1.48)	3.11 (1.63)	3.56 (1.67)
Moralinconsistency	3.56 (1.97)	4.09 (1.75)	4.63 (1.59)	4.23 (1.59)
Combinedinconsistency	3.33 (2.05)	4.54 (1.57)	3.90 (1.70)	4.27 (1.58)
Contrast: Moralinconsistency vs. attitudeinconsistency and combinedinconsistency	$t(394) = 1.86, p = .063,$ $d = 0.11 [-0.01, 0.23]$	$t(394) = -0.95, p = .345,$ $d = -0.06 [-0.18, 0.06]$	$t(394) = 5.62, p < .001,$ $d = 0.34 [0.22, 0.46]$	$t(394) = 1.58, p = .116,$ $d = 0.10 [-0.02, 0.22]$

Note: The contrast code was constructed such that attitudeinconsistency = -0.25 , combinedinconsistency = -0.25 , and moralinconsistency = $+0.50$.

and strong attitudeinconsistency incarnations. Interestingly, however, the strongest difference was on universalism: perceivers thought that Kevin would apply his beliefs to everybody (presumably including themselves) more in the moralinconsistency versus the two attitudeinconsistency conditions. No differences emerged for objectivity or emotionality. Attribution style had no main or interactive effects on any variables, $F_s < 1.09, p_s > .357$.

This finding may indicate that moral inconsistency is uniquely resistant to situational considerations in part because perceivers think that moral judgements imply universal claims. Indeed, among the three inconsistency conditions, only universalism was correlated with higher hypocrisy ratings, $r(298) = .28, p < .001$. Believing that Kevin felt his attitude was moral, $r(298) = .00, p = .950$; objective, $r(298) = .11, p = .050$; or emotional, $r(298) = .10, p = .087$, were each only weakly related to heightened hypocrisy judgements. Thus, among the various properties asserted about moral attitudes in past work, universalism is unique both in that other people espousing moral (vs. non-moral) opinions only increased the perception that the target's opinion was universal, and in that only universalism led people to judge the target as more hypocritical.

5 | GENERAL DISCUSSION

Despite a growing literature suggesting that moral attitudes may be distinct from other strongly held attitudes (Mueller & Skitka, 2018; Skitka, 2010; Skitka et al., 2005, 2021), there is limited work on whether people use this distinction when judging other people's attitudes. I examined this question with three experiments. In some respects, my experiments indicated that moral inconsistency was evaluated by perceivers in very similar ways from sufficiently defined strong attitudes. Hypocrisy from moral attitudes was evaluated as similarly hypocritical and similarly unlikeable as hypocrisy from attitudes that were committed and embedded (all experiments) and were judged to be similarly surprising and similarly reflective of inauthenticity, with a similar pattern of mediations emerging when testing if surprise and inauthenticity mediated between attitude strength and (low) likeabil-

ity. I accrued meaningful evidence for the null hypothesis across all relevant tests.

However, some interesting distinctions emerged in Experiment 3: although perceivers judged even a very strong (committed, embedded) attitude-inconsistent target as less hypocritical when using situational thinking, perceivers showed no such attributional shift for moral hypocrites. Thus, despite extensive evidence for moral attitudes being viewed as interchangeable with committed/embedded attitudes, a situational manipulation prompted a subtle but important difference in how these attitude strength qualities are appraised. Interestingly, moral attitudes also diverged in one other essential dimension: perceivers assumed moral attitudeholders to be implicitly universalizing their stances to everybody.

5.1 | Theoretical contributions

5.1.1 | Hypocrisy

Most obviously, my findings speak to the hypocrisy literature. First, I conceptually replicated past work (i.e., Jordan et al., 2017; Teeny et al., 2023) using materials and procedures that with some novelties (i.e., within-participant design; different attitude objects as discussed momentarily).

Second, unlike most hypocrisy studies, I deployed morally vacuous attitude objects. Nonetheless, I demonstrated pronounced hypocrisy and negative evaluation results even towards targets who expressed regular attitudes and then acted inconsistently. For instance, in Experiments 1 and 2, perceivers rated targets as approximately a standard deviation more hypocritical and less likeable when they acted inconsistently from an attitude, versus doing the same behaviour without such an attitude expression. Since most hypocrisy research concerns morally loaded topics like criminal behaviour (Efron et al., 2018; Laurent et al., 2014), capital punishment (Kreps et al., 2017), use of prohibited substances in sport (Jordan et al., 2017) and wearing animal fur (Laurent & Clark, 2019), the fact that pronounced hypocrisy judgements are rendered against targets who express not especially

strong opinions on taking one's dog to the pub or ordering takeout food (before doing the opposite) is noteworthy.

Third, I showed that perceivers treated moral attitude inconsistency as no more problematic than inconsistency from committed, embedded attitudes. This demonstrates that to a great extent it is the strength of a target's attitude, rather than its moral qualities per se, that may drive much of how perceivers are likely to evaluate a target. This helps to bring the hypocrisy literature, and person perception more broadly, closer in line with the attitude strength literature which has often indicated that morality is deeply connected with other kinds of attitude strength (e.g., Philipp-Muller et al., 2020).

5.1.2 | Moral 'versus' strong attitudes

The present findings may also speak to the ongoing debate about whether moral attitudes are distinct from (sufficiently defined) 'strong' attitudes. Interestingly, I found no evidence that perceivers expected moral attitudes to be believed more objective by targets, or associated with greater emotionality, although these are each qualities sometimes attributed to moral attitudes. At first, this may seem to contradict the work by Skitka and colleagues (2005) in which moral attitudes were felt by attitudeholders to be more objective and emotionally held. However, most research on moral attitudes concerns the self-rated properties of moral (vs. non-moral) opinions according to the opinionholders, whereas the present findings concern *perceivers' understanding* of what properties moral (vs. non-moral) opinions of other people have.

Additionally, most research on moral attitudes also compares moral attitudes to non-moral attitudes, rather than comparing moral attitudes across attitudes that are non-moral but are *strong for other reasons*. In this sense, one might consider the present findings to represent a very stringent test of whether people consider moral attitudes (of others) to be 'special'. Only universalism survived this stringent test in that it was uniquely elevated only for moral and not for committed/embedded attitudes. Furthermore, only universalism was substantially related to judging the target as hypocritical. This is reminiscent of Laurent and Clark's (2019) finding that hypocrisy was perceived as strongest when people imposed their moral attitudes on others before contradicting those attitudes.

Why might universalism (vs. objectivity and emotionality) be 'special' among the moral attitude properties in this domain in its association with hypocrisy? Past work has raised a false signalling of moral superiority account which suggests that hypocrites give a false impression of their values and future behaviours, and thereby '[shame] other people into changing their behaviour while the hypocrite carries on' (Jordan et al., 2017, p. 12; also see Laurent & Clark, 2019). A morally inconsistent person's perceived universalism may most strongly fuel this process because it implies that the target is attempting control over other people's actions – which is then judged manipulative if the target does not act consistently themselves. This may further explain why moral attitudes were resistant to perceivers forming situational attributions for targets' inconsistencies: situational attributions

require a perceiver to be motivated to adjust from trait inferences (Gawronski, 2004), and perceivers' motivation was undermined by feeling that the moral hypocrite engaged in manipulative false signalling.

One limitation of the present work is that I only measured universalism. However, universalism could be experimentally manipulated as a moderator to investigate its role as a mechanism of hypocritical inconsistencies (Spencer et al., 2005). For instance, participants could be presented with one target expressing a deeply held moral belief they do not apply to others, and another imposing the same moral belief on others. One could then test if perceivers are willing to use situational attributions for the first target (i.e., evaluating them less harshly when situational factors are weighed in) but maintain harsh judgements of the second target.

However, it is also possible that a property of moral attitudes other than what I measured drove this effect. For instance, moral (vs. non-moral) attitudes may be viewed as a bolder self-promotional claim of virtue (Jordan et al., 2017) or be associated with a greater sense of behavioural obligation (Kouchaki et al., 2018; Kreps et al., 2017; Sabucedo et al., 2018). Perceivers may hold such targets to these claims and obligations, reasoning that if a target's view is moral for the target, then there is no excuse for the target not to act consistently with it because attitude-consistent behaviour is a moral obligation rather than 'merely' a committed/embedded opinion. Addressing this question would further illuminate how people perceive and condemn hypocrisy.

FUNDING INFORMATION

This research was not funded by any awards body.

CONFLICT OF INTEREST STATEMENT

The author has no conflict of interest to disclose.

DATA AVAILABILITY STATEMENT

The data is openly available at <https://osf.io/whu9z/>.

ETHICS APPROVAL STATEMENT

All participants gave informed consent before participating in the research. The research was granted ethics clearance by a departmental ethics review board and conforms to the APA Code of Conduct. The research was conducted ethically, results are reported honestly, the submitted work is original and not (self-) plagiarized, and authorship reflects individual contributions.

TRANSPARENCY STATEMENT

All results are reported honestly, the studies were conducted ethically, and the submitted work is original. All data, variables, and codes are publicly available at <https://osf.io/whu9z/>.

ORCID

Thomas Ian Vaughan-Johnston  <https://orcid.org/0000-0002-4682-481X>

REFERENCES

- Abelson, R. P. (1988). Conviction. *American Psychologist*, 4, 267–275.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211.
- Alicke, M., Gordon, E., & Rose, D. (2013). Hypocrisy: what counts? *Philosophical Psychology*, 26(5), 673–701.
- Bailey, E. R., & Levy, A. (2022). Are you for real? Perceptions of authenticity are systematically biased and not accurate. *Psychological Science*, 33(5), 798–815.
- Bassili, J. N. (2008). Attitude strength. In W. D. Crano & R. Prislin (Eds.), *Attitudes and attitude change*. Psychology Press.
- Brandt, M. J., & Wetherell, G. A. (2012). What attitudes are moral attitudes? The case of attitude heritability. *Social Psychological and Personality Science*, 3(2), 172–179.
- Conner, M., & Armitage, C. J. (1998). Extending the theory of planned behavior: A review and avenues for further research. *Journal of Applied Social Psychology*, 28(15), 1429–1464.
- Crant, J. M. (1996). Doing more harm than good: When is impression management likely to evoke a negative response? *Journal of Applied Social Psychology*, 26(16), 1454–1471.
- Eastman, K. K. (1994). In the In eyes of the beholder: An attributional approach to ingratiation and organizational citizenship behavior. *Academy of Management Journal*, 37(5), 1379–1391.
- Effron, D. A., Markus, H. R., Jackman, L. M., Muramoto, Y., & Muluk, H. (2018). Hypocrisy and culture: Failing to practice what you preach receives harsher interpersonal reactions in independent (vs. interdependent) cultures. *Journal of Experimental Social Psychology*, 76, 371–384.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160.
- Fazio, R. H., & Zanna, M. P. (1981). Direct experience and attitude-behavior consistency. In *Advances in experimental social psychology* (Vol. 14, pp. 161–202). Academic Press.
- Garcia, A. C. (2019). *On liking and perceived authenticity* [unpublished undergraduate dissertation]. Texas A&M University. <https://oaktrust.library.tamu.edu/bitstream/handle/1969.1/175411/GARCIA-FINALTHESIS-2020.pdf?sequence=1>
- Gawronski, B. (2004). Theory-based bias correction in dispositional inference: The fundamental attribution error is dead, long live the correspondence bias. *European review of social psychology*, 15(1), 183–217.
- Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, 117(1), 21–38.
- Gilbert, D. T., Pelham, B. W., & Krull, D. S. (1988). On cognitive busyness: When person perceivers meet persons perceived. *Journal of Personality and Social Psychology*, 54(5), 733.
- Gross, S. J., & Niman, C. M. (1975). Attitude-behavior consistency: A review. *Public Opinion Quarterly*, 39(3), 358–368.
- Heider, F. (1958). *The psychology of interpersonal relations*. Wiley.
- Holland, R. W., Verplanken, B., & van Knippenberg, A. (2003). From repetition to conviction: Attitude accessibility as a determinant of attitude certainty. *Journal of Experimental Social Psychology*, 39, 594–601.
- Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.
- Jongman-Sereno, K. P., & Leary, M. R. (2020). Self-judgments of authenticity. *Self and Identity*, 19(1), 32–63.
- Jordan, J. J., Sommers, R., Bloom, P., & Rand, D. G. (2017). Why do we hate hypocrites? Evidence for a theory of false signaling. *Psychological Science*, 28(3), 356–368.
- Kan, M. P., & Fabrigar, L. R. (2017). Theory of planned behavior. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of personality and individual differences* (pp. 1–8). Springer.
- Kernis, M. H., & Goldman, B. M. (2006). A multicomponent conceptualization of authenticity: Theory and research. *Advances in Experimental Social Psychology*, 38, 283–357.
- Kifer, Y., Heller, D., Perunovic, W. Q. E., & Galinsky, A. D. (2013). The good life of the powerful: The experience of power and authenticity enhances subjective well-being. *Psychological Science*, 24(3), 280–288.
- Kouchaki, M., Smith, I. H., & Savani, K. (2018). Does deciding among morally relevant options feel like making a choice? How morality constrains people's sense of choice. *Journal of Personality and Social Psychology*, 115(5), 788.
- Kraus, M. W., Chen, S., & Keltner, D. (2011). The power to be me: Power elevates self-concept consistency and authenticity. *Journal of Experimental Social Psychology*, 47(5), 974–980.
- Kreps, T. A., Laurin, K., & Merritt, A. C. (2017). Hypocritical flip-flop, or courageous evolution? When leaders change their moral minds. *Journal of Personality and Social Psychology*, 113(5), 730.
- Krosnick, J. A., Boninger, D. S., Chuang, Y. C., Berent, M. K., & Carnot, C. G. (1993). Attitude strength: One construct or many related constructs? *Journal of Personality and Social Psychology*, 65(6), 1132.
- Krosnick, J. A., & Petty, R. E. (1995). Attitude strength: An overview. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences*. Lawrence Erlbaum Associates.
- LaPiere, R. T. (1934). Attitudes vs. actions. *Social Forces*, 13(2), 230–237.
- Laurent, S. M., & Clark, B. A. (2019). What makes hypocrisy? Folk definitions, attitude/behavior combinations, attitude strength, and private/public distinctions. *Basic and Applied Social Psychology*, 41(2), 104–121.
- Laurent, S. M., Clark, B. A., Walker, S., & Wiseman, K. D. (2014). Punishing hypocrisy: The roles of hypocrisy and moral emotions in deciding culpability and punishment of criminal and civil moral transgressors. *Cognition & Emotion*, 28(1), 59–83.
- Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.
- Liu, Y., & Perrewe, P. L. (2006). Are they for real? The interpersonal and intrapersonal outcomes of perceived authenticity. *International Journal of Work Organisation and Emotion*, 1(3), 204–214.
- Luttrell, A., Petty, R. E., Briñol, P., & Wagner, B. C. (2016). Making it moral: Merely labeling an attitude as moral increases its strength. *Journal of Experimental Social Psychology*, 65, 82–93.
- Luttrell, A., Philipp-Muller, A., & Petty, R. E. (2019). Challenging moral attitudes with moral messages. *Psychological Science*, 30(8), 1136–1150.
- Markowitz, D. M., Kouchaki, M., Gino, F., Hancock, J. T., & Boyd, R. L. (2023). Authentic first impressions relate to interpersonal, social, and entrepreneurial success. *Social Psychological and Personality Science*, 14(2), 107–116.
- Mueller, A. B., & Skitka, L. J. (2018). Liars, damned liars, and zealots: The effect of moral mandates on transgressive advocacy acceptance. *Social Psychological and Personality Science*, 9(6), 711–718.
- Newman, G. E. (2019). The psychology of authenticity. *Review of General Psychology*, 23(1), 8–18.
- Petty, R. E., & Krosnick, J. A. (1995). *Attitude strength: Antecedents and consequences*. Erlbaum.
- Philipp-Muller, A. Z., Wallace, L. E., & Wegener, D. T. (2020). Where does moral conviction fit?: A factor analytic approach examining antecedents to attitude strength. *Journal of Experimental Social Psychology*, 86, 103900.
- Pomerantz, E. M., Chaiken, S., & Tordesillas, R. S. (1995). Attitude strength and resistance processes. *Journal of Personality and Social Psychology*, 69(3), 408.
- Portal, S., Abratt, R., & Bendixen, M. (2019). The role of brand authenticity in developing brand trust. *Journal of Strategic Marketing*, 27(8), 714–729.
- Prislin, R. (1996). Attitude stability and attitude strength: One is enough to make it stable. *European Journal of Social Psychology*, 26, 447–477.
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing <https://www.R-project.org/>

- Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In *Advances in experimental social psychology* (Vol. 10, pp. 173–220). Academic Press.
- Ross, L. D. (2001). Getting down to fundamentals: Lay dispositionism and the attributions of psychologists. *Psychological Inquiry*, 12(1), 37–40.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36.
- Sabucedo, J. M., Dono, M., Alzate, M., & Seoane, G. (2018). The importance of protesters' morals: Moral obligation as a key variable to understand collective action. *Frontiers in Psychology*, 9, 418.
- Sezer, O., Gino, F., & Norton, M. I. (2018). Humblebragging: A distinct—and ineffective—self-presentation strategy. *Journal of Personality and Social Psychology*, 114(1), 52.
- Sheldon, K. M., Ryan, R. M., Rawsthorne, L. J., & Ilardi, B. (1997). Trait self and true self: Cross-role variation in the Big-Five personality traits and its relations with psychological authenticity and subjective well-being. *Journal of Personality and Social Psychology*, 73(6), 1380–1393.
- Skitka, L. J. (2010). The psychology of moral conviction. *Social and Personality Psychology Compass*, 4(4), 267–281.
- Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral conviction: Another contributor to attitude strength or something more? *Journal of Personality and Social Psychology*, 88(6), 895–917.
- Skitka, L. J., Hanson, B. E., Morgan, G. S., & Wisneski, D. C. (2021). The psychology of moral conviction. *Annual Review of Psychology*, 72, 347–366.
- Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology*, 89(6), 845.
- Sutton, A. (2020). Living the good life: A meta-analysis of authenticity, well-being and engagement. *Personality and Individual Differences*, 153, 109645.
- Teeny, J. D., Lanzalotta, J. V., & Petty, R. E. (2023). Understanding the magnitude of hypocrisy in moral contradictions: The role of surprise at violating strong attitudes. *Personality and Social Psychology Bulletin*, <https://journals.sagepub.com/doi/epub/10.1177/01461672231177773>
- Zanna, M. P., Olson, J. M., & Fazio, R. H. (1980). Attitude–behavior consistency: An individual difference perspective. *Journal of Personality and Social Psychology*, 38(3), 432.
- Zigler, C. K., & Ye, F. (2019). A comparison of multilevel mediation modeling methods: recommendations for applied researchers. *Multivariate Behavioral Research*, 54(3), 338–359.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Vaughan-Johnston, T. I. (2024). Hypocrisy judgements are affected by target attitude strength and attitude moralization. *European Journal of Social Psychology*, 54, 397–414. <https://doi.org/10.1002/ejsp.3018>