Check for updates

Empirical Research Paper

"If Immoral Then Unable": Asymmetric Generalizations in Social Judgment

Personality and Social Psychology Bulletin
1–16
© 2026 by the Society for Personality and Social Psychology, Inc
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/01461672251410724
journals.sagepub.com/home/pspb
Sage

Simone Mattavelli[1], Marco Brambilla[1], Alex Koch[2], and Marcos Dono[3]

Abstract

Morality and ability are two key dimensions of social judgment. Across four experiments (total $N = 1{,}418$, three preregistered), we examined how information about one dimension shapes impressions in the other. In Experiment 1, participants generated positive and negative behaviors related to either morality or ability and then evaluated each behavior on the other dimension. Negative moral behaviors led to stronger inferences of low ability than negative ability behaviors led to inferences of immorality (i.e., asymmetric Horn effect). No asymmetry emerged for positive behaviors (i.e., symmetric Halo effects). Experiments 2a and 2b confirmed the asymmetric Horn effect and showed it was stronger for extreme versus moderate negative behaviors. Experiment 3 showed that immoral behaviors elicited more perceived threat than unable behaviors, which partly explained the asymmetric horn effect. These findings complement and extend prior models of impression formation by highlighting the primacy of morality in influencing judgments on other fundamental content dimensions.

Keywords

morality, horn effect, halo effect, impression formation, social judgment

Received: June 15, 2025; revision accepted: December 10, 2025

Introduction

Imagine two surgeons: one who deliberately recommends unnecessary operations, prioritizing financial gain over patient well-being, and another who makes critical errors during vital procedures due to a lack of skill. How do we perceive them? Does the first surgeon's unethical behavior lead you to question their medical ability? Conversely, does the second surgeon's incompetence prompt doubts about their moral integrity? More generally, does information about morality exert a stronger influence on perceptions of ability than information about ability exerts on perceptions of morality? Understanding this comparative impact of morality and ability highlights how reputational damage or enhancement spreads across evaluative domains and can inform strategies for impression management. Building on the classic Halo and Horn effects (e.g., Thorndike, 1920), we examine whether moral information (e.g., dishonesty) is more likely to influence judgments of ability than the reverse.

Morality and Ability in Social Judgment

A long tradition of research has shown that two broad social-evaluative dimensions—warmth and competence—shape emotional and behavioral reactions toward other individuals and groups (for a review, Abele et al., 2021). The "warmth" dimension (also called horizontal dimension) pertains to benevolence in social relations and involves qualities such as trustworthiness, cooperation, likability, friendliness, and honesty (Abele & Wojciszke, 2007, 2014; Brambilla & Leach, 2014; Ellemers et al., 2013; Leach et al., 2007; Wojciszke et al., 2011). The "competence" dimension (also called vertical dimension) refers to the power to achieve one's goals effectively and encompasses qualities such as assertiveness, confidence, ability, and intelligence (Abele & Wojciszke, 2007; Fiske et al., 2002; Wojciszke, 2005). Whereas warmth shapes impressions by revealing whether individuals' intentions toward us are beneficial or harmful, competence signifies whether an individual possesses the capability to enact their intentions toward us (Fiske et al., 2007).

In the last decades, researchers have proposed that both warmth and competence each include two separable

[1]University of Milano-Bicocca, Italy
[2]University of Chicago Booth School of Business, IL, USA
[3]University of Santiago de Compostela, Spain

Corresponding Author:
Simone Mattavelli, Department of Psychology, University of Milano-Bicocca, 1, Piazza dell'Ateneo Nuovo, Milan 20126, Italy.
Email: simone.mattavelli@unimib.it

components. Warmth encompasses the facets of sociability and morality (Brambilla & Leach, 2014; Brambilla et al., 2021; Leach et al., 2007; see also Abele et al., 2021; Koch et al., 2021; Yzerbyt et al., 2025). Sociability refers to individuals' tendencies to affiliate and form connections, whereas morality concerns perceptions of a person's virtue (Brambilla & Leach, 2014; Goodwin, 2015). Competence is similarly understood to include ability and assertiveness as two facets (Abele et al., 2021). Ability refers to a person's intelligence and skillfulness, whereas assertiveness stands out as a motivational facet of competence that reflects self-confidence and determination (Abele et al., 2016).

Among these facets, morality is more critical than sociability in evaluating warmth and in forming impressions (see Brambilla et al., 2021). Similarly, ability is typically regarded as more essential than assertiveness in evaluating competence such that the terms "ability" and "competence" have often been used interchangeably (Chen et al., 2014; De Bruin & Van Lange, 1999, 2000; Goodwin et al., 2014). Together, these distinctions motivate our focus on morality and ability as the most central facets of the two dimensions in social judgment (see Ellemers, 2017).

The Importance of Morality and Ability in Social Evaluation

Notwithstanding the importance of both morality and ability in social judgment, one widely investigated question pertains to the relative weight that each dimension has in shaping overall impressions about other individuals and groups. To determine the primacy of one dimension over the other, researchers considered multiple criteria (Abele et al., 2021; Nicolas et al., 2022). For instance, the global "positive" versus "negative" impression of an individual is influenced by moral traits more than by non-moral traits (Goodwin et al., 2014). Moral traits are also dominant in shaping one's self-concept (Christy & Schlegel, 2024; Strohminger & Nichols, 2014), and people feel that they know more about the true selves of highly moral individuals, compared to those distinguished mainly by non-moral characteristics (Christy et al., 2017). Reinforcing these findings, the moral primacy model of impression development (Brambilla et al., 2021) posited and demonstrated that throughout the multiple stages of impression development, comprising first impressions and their later updating, information about the target's morality is more influential than information about their competence (ability) or sociability (see Brambilla et al., 2019; Luttrell et al., 2022).

A growing body of work has also shown that morality is key in shaping group perception. Morality represents the strongest basis for group-related emotions and thoughts, such as pride in, and identification with, the group (Leach et al., 2007). Similarly, individuals anticipate receiving ingroup respect when adhering to morality-related norms (Pagliaro et al., 2011), indicating a specific concern for morality information when considering one's social identity and standing within the group one belongs to. Moral qualities are also relevant to impressions of outgroups. Indeed, global impression of the outgroup is most affected by its perceived morality as immoral outgroup members are perceived as posing a real and concrete danger to the in-group's safety (Brambilla et al., 2012, 2013).

Taken together, previous empirical findings converge to the idea that social perceivers' impressions and overall judgments are more influenced by morality than ability. However, common to these reviewed studies is the tendency to assess the relative importance of morality and ability by measuring their impact on an external criterion, including overall impressions and behavioral outcomes. A less explored perspective is how information about one dimension influences judgments on the other dimension. Whereas prior research has documented morality's primacy in shaping *overall impressions* of other individuals and groups (Brambilla et al., 2021; Goodwin et al., 2014), it remains unclear whether, and under which conditions, moral information exerts a stronger influence on perceptions of ability than the reverse. Here, we addresses this gap by systematically investigating cross-dimensional generalization between morality and ability, a domain in which the comparative power of morality and ability has not been thoroughly explored.

Halo and Horn: On the Bidirectional Influence Between Morality and Ability

Research on social judgment presents conflicting views on the relationship between key dimensions—morality and ability (or competence). Some studies suggest these dimensions are independent (Leach et al., 2007), while others find positive (Abele, 2003; Abele & Wojciszke, 2007; Durante et al., 2017; Koch et al., 2024), negative (Yzerbyt, 2016, 2018), curvilinear (Imhoff & Koch, 2017; Koch et al., 2016), or even variable relationships (Koch et al., 2020; Roberts & Koch, 2025; Stolier et al., 2020). However, a common limitation in these studies is their reliance on correlational methods, leaving open the question of causality.

The present research focuses less on the independence versus interdependence of these dimensions and more on their comparative, cross-dimensional influence. Specifically, we ask whether information about morality has a stronger effect on perceptions of ability than information about ability has on perceptions of morality. This inquiry aligns with classical Halo and Horn effects (Thorndike, 1920), where positive traits in one observed domain can enhance inferences in another non-observed domain (i.e., Halo effect) and negative traits can reduce them (i.e., Horn effect).

Prior work provides some relevant insights. For example, "compensation effects" have been observed in group evaluations, where low ratings in one dimension are offset by higher ratings in another (Judd et al., 2005; Yzerbyt et al.,

2005, 2008). However, this tendency has been documented primarily in intergroup contexts, where compensation reflects the existence of an accepted asymmetry between social groups that is functional for regulating intergroup relations (Cambon et al. 2015; Yzerbyt & Cambon, 2017). Because these motives are less salient when judging individuals, there is less reason to expect compensatory inferences to emerge in person perception. Consistent with this view, Gräf and Unkelbach (2016) found that positive traits produce cross-domain Halo effects. For instance, honesty (a morality trait) enhanced perceptions of industriousness (an ability trait). In contrast, negative traits only produced within-domain Horn effects, such as habitual lying leading to perceptions of selfishness, rather than generalizing across dimensions (i.e., to a trait that captures some aspect of inability). Moreover, Stellar and Willer (2018) explored the Horn effect in a unidirectional manner—how immoral behaviors influence perceptions of (in)ability. They found that negative moral behaviors (e.g., cheating on a spouse) generalized to negative judgments on the ability domain, demonstrating that perceived immorality can undermine evaluations of ability.

These partially inconsistent findings leave key questions unanswered. Namely, how reliable is cross-dimensional generalization for positive versus negative traits? More importantly, does generalization occur more strongly from morality to ability than from ability to morality? And is this asymmetry influenced by the valence of the behaviors? By directly testing both the asymmetry and the role of valence, we can determine whether the hypothesized asymmetric effect is conditional on negative versus positive information, identifying valence as a boundary condition. This approach not only clarifies which dimension is functionally more central in guiding impression formation but also provides insight into when and how reputational damage or enhancement spreads across evaluative domains.

The Asymmetric Generalization Hypothesis

We advance the *Asymmetric Generalization Hypothesis*, whereby morality-relevant information is expected to affect ability more than the reverse. We test this hypothesis by considering the role of relevant moderators, such as valence, extremeness, and self-relevance. Importantly, the role of such moderators in qualifying the hypothesized asymmetric generalization allows us testing different prediction made by alternative accounts of impression formation. For instance, a cognitive-ecological *account* suggests that the informativeness of a behavior depends on its frequency/rarity in the social environment (Rusconi et al., 2020; Skowronski & Carlston, 1987). Across most social contexts, highly moral and low ability behaviors tend to be more frequent than immoral acts and displays of high ability. Consequently, immorality and high ability are seen as more revelatory of person's true character (Fiske, 1980; Reeder & Brewer, 1979). This explanation would predict an interactive effect between dimension and valence: not only a stronger negative generalization from immorality to inability (vs. from inability to immorality), but also a stronger positive generalization from ability to morality (vs. from morality to ability).

However, people may generalize impressions of others' (im)morality and (in)ability in a self-serving way that promises the fulfillment of their fundamental social motives, including protecting the self from social threats (Cook et al., 2021; Kenrick et al., 2010). Self-serving generalizations would operate via an affective-motivational process. On the positive side, the perceived morality of a target person (e.g., their loyalty) likely serves the perceiver (and others) better if the target turns out to be able (e.g., they are beneficial; Brambilla et al., 2021; Fiske et al., 2007). The reverse is also likely to be true. As a result, the self-interest account predicts that direction of generalization (from morality to ability vs. the reverse) does not moderate the strength of the Halo effect. The story is different for the dark side. Assuming that immoral individuals are also unable might serve the self-protective function of reducing the perceived threat from their malevolent intentions. Indeed, situations in which others seem both immoral and competent are highly undesirable and ought to be avoided because they can be expected to cause harm and suffering (Carrier et al., 2019; Landy et al., 2016, Roberts & Koch, 2025). In contrast, seeing someone as unable does not make a generalization to their immorality particularly self-serving because inability limits the harm and suffering that immorality can tolerate or even motivate. Thus, the affective-motivational account predicts the Horn asymmetry that perceived immorality is more likely to generalize to impressions of low ability, compared to the likelihood of inferring immorality from low ability.

In summary, cognitive-ecological and affective-motivational theorizing converges on the asymmetric Horn hypothesis that people infer inability from perceived immorality more than the reverse. The cognitive-ecological account predicts an asymmetric Halo effect, with a stronger inference of morality from perceived ability (vs. the reverse). In contrast, the affective-motivational account predicts symmetric generalization between perceived morality and ability.

Overview of the Present Research

Our research explores generalization in social judgment from (im)morality to (in)ability and vice versa. We conducted four experiments (three preregistered). In Experiment 1, participants generated positive and negative behaviors relevant to either morality or ability and then evaluated these behaviors on the alternative dimension. Experiments 2a and 2b focused exclusively on negative behaviors (i.e., the Horn effect), crossing manipulations of behavior dimension and extremeness, a potential confound variable introduced by the specific task adopted in Experiment 1. In a highly powered test, Experiment 3 examined the Horn effect, testing whether

immoral behaviors generalize to perceived inability more than vice versa. Experiment 3 also explored the roles of behavior extremeness and perceived threat, with self-relevance as a moderator. Immoral (vs. low ability) behaviors may elicit stronger threat perceptions, and thereby a larger Horn effect.

We did not substantiate an asymmetric Halo effect. But across experiments, our findings consistently support an asymmetric Horn effect. Immoral behaviors more strongly influenced perceptions of inability than inability behaviors influenced perceptions of immorality. Furthermore, we found that behavior extremeness amplified this effect: the more severe the observed immoral and low ability behaviors, the stronger the asymmetric Horn effect. A key mechanism underlying this asymmetry may be perceived threat, consistent with the affective-motivational account. Participants rated immoral (vs. unable) behaviors as more threatening, which statistically mediated people's greater inference of inability from behavioral immorality (vs. the reverse inference of immorality from behavioral inability).

Open Science

We preregistered all experiments but Experiment 1 (Supplemental Material Experiment 1b: https://osf.io/5yxs6/; Experiment 2a: https://osf.io/qa758/; Experiment 2b: https://osf.io/muhg3/; Experiment 3: https://osf.io/6mgfs/).[1] The materials, data, and analysis code for the experiments are available through the website of the Open Science Foundation (https://osf.io/rfa76/). We report all manipulations and measures used in the experiments. All experiments received formal ethics approval by the university that funded the present research.

Experiment 1

Participants generated positive and negative behaviors related to morality and ability. We examined whether these behaviors would elicit a Halo or Horn effect, which means a generalization to the same valence on the alternative dimension. Because the task involved self-generated behaviors, we anticipated the chance that some domains (e.g., morality) would naturally elicit more extreme exemplars than others (e.g., ability). Hence, we measured extremeness alongside generalization.

Method

We employed a 2 (Behavior domain: morality vs. ability) × 2 (Behavior valence: positive vs. negative) mixed design, with the second factor manipulated within participants. We measured how each behavior influenced judgments on the alternative dimensions.

Sensitivity Analysis. Using the software GPower 3.1 (Faul et al., 2009), we conducted a sensitivity analysis on the within-between interaction effect between behavior domain and valence. With alpha = .05, and power (1-beta) = .90, the analysis indicates that with a final sample of 228 valid-data participants, we could detect an effect as small as $\eta^2 = .025$.

Participants and Procedure. We recruited 302 participants (134 females, 166 males, 2 unspecified, $M_{age} = 38.07$ years old, $SD_{age} = 11.27$ years old) online to participate in a study about impression formation via Prolific Academic. After they had agreed to participate and given informed consent, participants were to generate five moral (or able) and five immoral (or unable) behaviors. Specifically, in this behavior generation phase, participants were first presented with five empty text boxes to be filled with five behaviors that they deemed as "intelligent/skilled/competent" (in the ability condition) or "honest/moral/trustworthy" (in the morality condition); and then with other five text boxes to be filled with five behaviors that they deemed as "unintelligent/unskilled/incompetent" or "dishonest/immoral/untrustworthy." Then, in the judgment phase, the 10 self-generated behaviors were presented again to the participants; their task was to rate them in the alternative dimension. Specifically, participants who generated ability-relevant behaviors rated each behavior on morality (i.e., from dishonest/immoral/untrustworthy to honest/moral/trustworthy), while participants who generated morality-relevant behaviors rated each behavior on ability (i.e., from unintelligent/unskilled/incompetent to intelligent/skilled/competent, see Brambilla et al., 2021). Next, all participants rated the self-generated behaviors for valence (i.e., from bad/dislikeable/negative to good/likable/positive). Both the morality/ability and the valence ratings consisted of an 11-point Likert scale ranging from −5 to +5.

To assess the validity of the generated behaviors, we asked two independent judges to make a dichotomous rating (i.e., 0 vs. 1) for each behavior. We assigned a score of 0 for text entries that did not represent actual behaviors (e.g., random text, dispositional traits, etc.); a score of 1 indicated any valid behaviors. In cases of disagreement between the scores assigned by each judge, the two judges collaboratively reviewed each target behavior and reached a consensus score. Eventually, all participants received an overall validity score ranging from 0 to 5, for both self-generated positive and negative behaviors. We excluded participants scoring below 2 on either positive or negative behaviors from the analyses (N = 74). This led to a final sample size of 228 participants.

Results

For positive behaviors, the original ratings that participants provided when evaluating each behavior on the alternative

dimension reflected the Halo scores. For negative behaviors, before analyzing the data, we flipped the scale on which the participants rated each behavior on the alternative dimension. This resulted in higher scores for stronger Horn, allowing a direct comparison between the generalization of positive behaviors (i.e., the Halo effect) versus negative behaviors (i.e., the Horn effect). Thus, for both positive and negative behaviors, a score of 5 indicated the strongest generalization effect possible. We used an equivalent approach to calculate a valence extremeness score. Again, a valence extremeness score of 5 indicated the most extreme valence ratings.

Generalization. We analyzed the data through a 2 (Behavior domain: morality vs. ability) × 2 (Behavior valence: positive vs. negative) mixed ANOVA. We found a main effect of behavior domain, $F(1,226) = 8.09$, $p = .005$, $\eta^2 = .025$, indicating an overall stronger generalization effect when generalizing from morality to ability ($M = 2.99$, $SD = 1.73$) compared to when generalizing from ability to morality ($M = 2.42$, $SD = 1.85$). There was a main effect of behavior valence, $F(1,226) = 32.18$, $p < .001$, $\eta^2 = .033$, indicating a stronger generalization effect for positive ($M = 3.04$, $SD = 1.54$) as opposed to negative ($M = 2.38$, $SD = 2.00$) behaviors. The interaction effect was significant, $F(1,226) = 22.11$, $p < .001$, $\eta^2 = .023$. Decomposing this interaction revealed that the effect of behavior domain (i.e., stronger generalization effect in morality-relevant behaviors) was significant for negative behaviors (i.e., the Horn effect), $t(226) = -4.36$, $p < .001$, $d = 0.58$, but not for positive behaviors (i.e., the Halo effect), $t(226) = 0.12$, $p = .906$, $d = 0.02$ (see Figure 1 with original positive and negative scores).

Extremeness. We conducted the same analysis using extremeness as the dependent variable. We found a main effect of behavior domain, $F(1,226) = 31.01$, $p < .001$, $\eta^2 = .088$, indicating that the behaviors generated in the moral domain were more extreme ($M = 4.15$, $SD = 1.25$) than those generated in the ability domain ($M = 3.25$, $SD = 1.62$). There was a main effect of behavior valence, $F(1,226) = 38.72$, $p < .001$, $\eta^2 = .035$, indicating higher extremeness for positive ($M = 3.99$, $SD = 1.17$) than negative ($M = 3.43$, $SD = 1.75$) behaviors. The interaction effect was significant, $F(1,226) = 30.12$, $p < .001$, $\eta^2 = .027$. The effect of behavior domain on valence extremeness (i.e., stronger extremeness for behaviors generated in the morality domain as opposed to the ability domain) was stronger for negative behaviors, $t(192) = 6.53$, $p < .001$, $d = 0.87$, than for positive behaviors, $t(222) = 2.62$, $p = .009$, $d = 0.35$.

Discussion

The results of Experiment 1 suggest that people have a higher tendency to generalize from perceived morality to ability than the other way around (i.e., a generalization asymmetry).

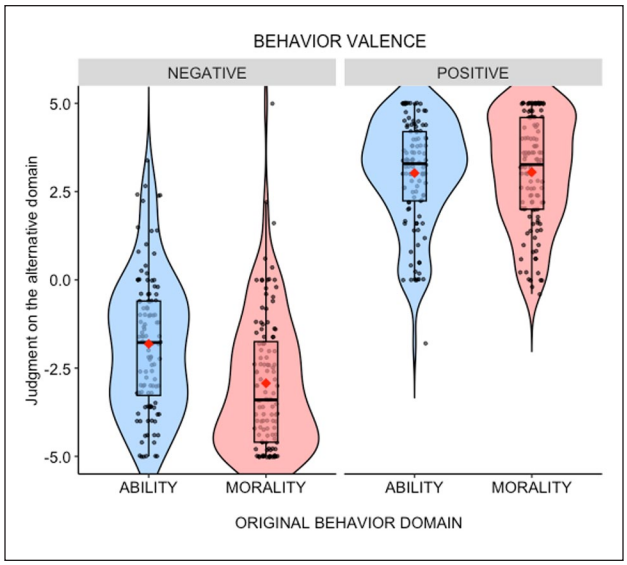

Figure 1. Effect of original behavior domain on between-dimension generalization for positive and negative behaviors—Experiment 1.

Note. Boxplots are the interquartile range, and the bars represent the median; the red diamonds represent the mean. The distributions are the kernel probability density of the data.

This asymmetry emerged for negative, but not for positive behaviors. In other words, an immoral behavior produced greater impact on judgments of inability as opposed to what an unable behavior did on judgments of immorality. The same interaction emerged on extremeness. A higher extremeness of the morality-relevant (vs. ability-relevant) behaviors was stronger for the negative than for the positive self-generated behaviors. Thus, the asymmetric Horn (but not Halo) effect that we observed may be due to the asymmetric extremeness of the negative (but not so much positive) behaviors that the participants had generated themselves. Experiments 2a and 2b aimed to rule out this alternative explanation, among other aims.

Experiments 2a-2b

We aimed to replicate the larger generalization effect from immoral behaviors onto judgments of ability (vs. from unable behaviors onto judgments of immorality), and to examine whether this asymmetric Horn effect is moderated by behavior extremeness (see also Experiment 1b, reported in the Supplemental Materials, which preliminary addressed this question). Because Experiment 1 found a symmetric Halo effect (for positive behaviors), in Experiments 2a and 2b we focused exclusively on negative behaviors, that is, the Horn effect. In addition to comparing the size of the Horn effect between the two behavior domains of immorality and inability, we manipulated the extremeness of the negative behaviors (mildly negative vs. highly negative). Moreover, while Experiment 1 asked participants to generalize judgments

across behaviors (e.g., the extent to which an immoral behavior was also perceived as lacking ability), Experiments 2a and 2b assessed generalization at the person level (e.g., the extent to which a person who engaged in an immoral behavior was also seen as unable). To do so, we presented participants with faces paired with a single negative behavior. The use of facial stimuli was intended to evoke judgments of concrete individuals rather than abstract acts, thereby moving closer to everyday impression formation.

The two experiments were largely identical, except for an additional measure introduced in Experiment 2b. In Experiment 2b, we also examined whether immoral and unable behaviors differ in their generalization within the same superordinate dimension. Specifically, we measured generalization to person-level low sociability for immoral behaviors (within the horizontal dimension), and to person-level low assertiveness for unable behaviors (within the vertical dimension; Abele et al., 2021; Koch et al., 2021; Yzerbyt et al., 2025). Prior research shows that the subordinate facets within each superordinate dimension exhibit high and comparably strong conceptual and empirical associations (Chawinwit & Koch, 2025; Koch et al., 2024). Because neither pair of facets (morality-sociability vs. ability-assertiveness) has an inherent associative advantage, testing both between- and within-dimension generalization tests the broader reach of morality-related information: if immoral behaviors generalize more strongly to low sociability than unable behaviors generalize to low assertiveness, this would indicate that negative moral information exerts a broader evaluative influence on person impressions than negative ability information.

Method

We employed a 2 (Behavior domain: immorality vs. inability) × 2 (Behavior extremeness: mildly negative vs. highly negative) mixed design, with the latter factor manipulated within participants. The main dependent variable was the generalization of judgments onto the alternative dimension (i.e., the Horn effect). In Experiment 2b, we also measured generalization within the same superordinate dimension: low sociability (for immoral behaviors) and low assertiveness (for unable behaviors).

Sensitivity Analysis. In the preregistration protocols, we reported separate power analyses for the two studies. Because the article presents the studies in aggregate, we instead report a sensitivity analysis, which we consider more appropriate and consistent with the other studies. Using the software GPower 3.1, we conducted two sensitivity analyses on the within-between interaction between behavior domain and extremeness. With alpha = .05 and power (1-beta) = .90, the first analysis indicates that with a final sample of 487 valid-data participants (pooling across Experiments 2a and 2b), we could detect a difference in the between-dimension generalization effect as small as $\eta^2 = .014$. The second analysis indicates that with a final sample of 308 valid-data participants (Experiment 2b only), we could detect a difference in the within-dimension generalization effect as small as $\eta^2 = .022$.

Participants and Procedure. Two hundred and six Prolific workers participated in Experiment 2a (116 females, 84 males, 5 unspecified, M_{age} = 39.58 years old, SD_{age} = 12.35 years old) and 379 Prolific workers participated in Experiment 2b (179 females, 186 males, 14 unspecified, M_{age} = 41.38, SD_{age} = 11.45). After providing informed consent, participants were randomly assigned to one of the two behavior domain conditions. In the first part of the study, we presented participants with a list of 45 negative behaviors (29 from Experiment 1 plus 16 newly generated behaviors), relevant for either morality (e.g., "Spreading false information," "Stealing from stores," and "Cheating when playing cards") or ability (e.g., "Missing an important deadline," "Falling asleep in a meeting," and "Asking a junior colleague to solve a problem"). Participants rated each behavior on an 11-point Likert scale (−5 = *extremely negative*; +5 = *extremely positive*). For each participant, our survey code selected two behaviors that received a highly negative score (−4) and two behaviors that received a mildly negative score (−2). If the number of behaviors receiving a score of −4 and −2 was lower than two, the script picked behaviors receiving a score of −5 (in place of −4, highly negative) and −1 (in place of −2, mildly negative). We excluded from the analyses all participants who did not assign these scores to at least one behavior (Experiment 2a: $N = 27$; Experiment 2b: $N = 71$). This led to a final sample of 179 participants in Experiment 2a and 308 in Experiment 2b.

Next, on four randomly ordered survey pages participants saw the faces of four individuals from the Chicago Face Database (Ma et al., 2015), each accompanied by one of the four behaviors (i.e., a mildly or highly negative one). Facial stimuli were white male identities selected to fall relatively close to the midpoint of the 1 to 7 trustworthiness scale (*M*s = 3.32–3.81 across the four faces used).[2] We asked the participants to evaluate each individual on the same morality Likert scale (that ranged from −5 to +5) as in Experiment 1 if the displayed behavior signaled inability, and on the same ability Likert scale as in Experiment 1 if the displayed behavior signaled immorality.[3] In Experiment 2b, after measuring between-dimension generalization, we also measured within-dimension generalization by asking the participants to judge each target individual on a second facet of the same superordinate dimension. That is, people rated the sociability of each individual (on a scale that ranged from unsociable/unfriendly/unhelpful to sociable/friendly/helpful) if the individual's behavior signaled immorality, and they rated the individual's assertiveness (on a scale that ranged from unassertive/

passive/dependent to assertive/ambitious/independent) if their behavior signaled inability.

Results

To test between-dimension (Horn) generalization, we aggregated the data from Experiments 2a and 2b. We conducted a 2 (Behavior domain: immorality vs. ability) × 2 (Behavior extremeness: highly negative vs. mildly negative) mixed ANOVA. The dependent variable was person-level impressions of morality (if the behavior domain was inability) or ability (if the behavior domain was immorality). We tested within-dimension generalization through a mixed ANOVA with the same factors, but here the dependent variable was person-level impressions of assertiveness (if the behavior domain was inability) or sociability (if the behavior domain was immorality). Analyses considering the two studies separately are reported in the Supplemental Materials.

Horn Generalization Between Dimensions. The effect of behavior domain was significant, $F(1,485) = 5.54$, $p = .019$, $\eta^2 = .008$, indicating a stronger tendency to judge the target as negative on ability when associated with an immoral behavior ($M = -1.61$, $SD = 1.78$), as opposed to judging the target as negative on morality when associated with an unable behavior ($M = -1.31$, $SD = 1.46$). There was a main effect of behavior extremeness, $F(1,485) = 102.49$, $p < .001$, $\eta^2 = 0.044$, indicating a stronger generalization effect for highly negative behaviors ($M = -1.82$, $SD = 1.48$) as opposed to mildly negative behaviors ($M = -1.13$, $SD = 1.74$). The interaction effect was not significant, $F(1,485) = 1.92$, $p = .166$, $\eta^2 = .001$ (see Figure 2).

Horn Generalization Within Dimensions. The effect of behavior domain only approached significance, $F(1,306) = 3.67$, $p = .056$, $\eta^2 = .008$. There was a main effect of behavior extremeness, $F(1,306) = 52.54$, $p < .001$, $\eta^2 = .044$, indicating a stronger effect for highly negative ($M = -2.45$, $SD = 1.61$) as opposed to mildly negative behaviors ($M = -1.78$, $SD = 1.54$). The interaction effect was significant, $F(1, 306) = 4.79$, $p = .029$, $\eta^2 = .004$. Decomposing this interaction revealed that the effect of behavior domain (i.e., stronger generalization effect for immoral behaviors on perceived sociability as opposed to unable behaviors on perceived assertiveness) was significant for highly negative behaviors, $t(306) = -2.80$, $p = .005$, $d = 0.31$, but not for mildly negative ones, $t(300) = 0.61$, $p = .61$, $d = 0.06$ (see Figure 3).

Discussion

Experiments 2a and 2b showed a larger Horn effect for immoral behaviors (generalizing onto ability judgments) compared to unable behaviors (generalizing onto morality judgments), consistent with previous findings. Extending Experiment 1's findings, Experiments 2a and 2b showed this asymmetric Horn effect when people generalized from an individual's behavior to their traits (vs. the within-behavior

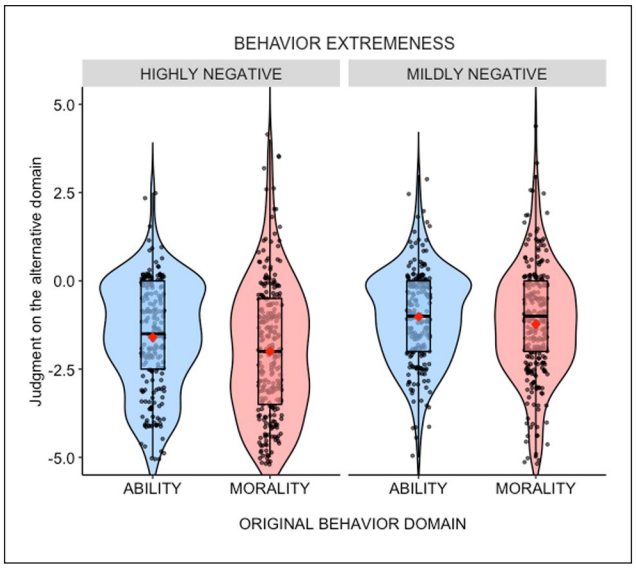

Figure 2. Effect of original behavior domain on between-dimension generalization for highly negative and mildly negative behaviors—Experiments 2a and 2b.
Note. Boxplots are the interquartile range, and the bars represent the median; the red diamonds represent the mean. The distributions are the kernel probability density of the data.

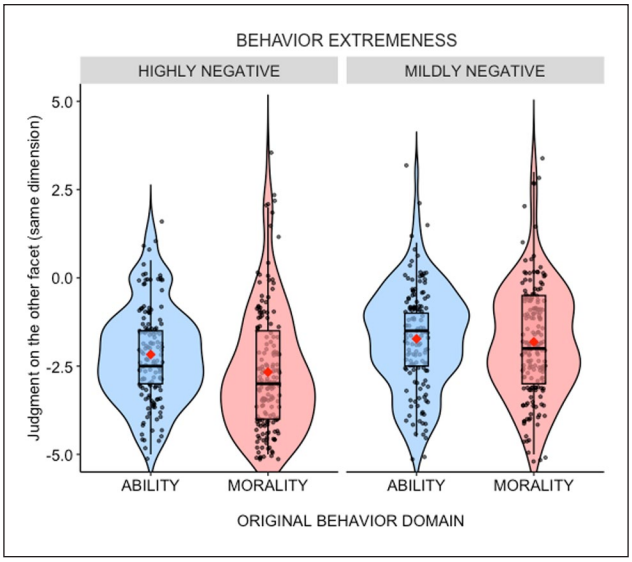

Figure 3. Effect of original behavior domain on within-dimension generalization (i.e., observed immorality generalizing to judgments of sociability, and observed inability generalizing judgments of assertiveness) for highly negative and mildly negative behaviors—Experiment 2b.
Note. Boxplots are the interquartile range, and the bars represent the median; the red diamonds represent the mean. The distributions are the kernel probability density of the data.

generalization in Experiment 1). In addition, the results of Experiment 2b showed an asymmetry in Horn generalization within dimensions. Highly immoral behaviors more strongly spilled over to impressions of low sociability, compared to the spillover from highly unable behaviors to perceiving the behaving individual as low on assertiveness. This asymmetry did not emerge for the mildly negative behaviors.

Experiment 3

The aim of Experiment 3 was twofold. First, we sought to conduct a highly powered test of the asymmetric Horn effect and the moderating role played by extremeness. Specifically, we examined whether the asymmetric Horn effect is larger for extremely negative behaviors compared to mildly negative ones, an interaction that emerged as significant in Experiment 2b when we analyzed Horn generalization within dimensions. Second, we tested the affective-motivational account outlined in the introduction. Immoral behaviors, compared to unable ones, may elicit stronger perceptions of threat, which may ultimately amplify the Horn effect as the perceiver can alleviate or even neutralize the perceived threat by perceiving the behaving individual as unable. We explore this in two ways. First, we introduced a measure of perceived threat to test (a) that immoral behaviors are perceived as more threatening than unable behaviors, and (b) that the magnitude of the asymmetry in the Horn effect correlates with differences in perceived threat. Second, we experimentally manipulated the self-relevance of the target individuals. In the self-irrelevant condition, participants judged individuals showing immoral or unable behaviors under the assumption that they would never encounter them in real life. In the self-relevant condition, participants judged the individuals while imagining that they could interact with them later in the study. Because self-relevance should heighten the perception of personal threat, we hypothesized that the asymmetric Horn effect would be more pronounced in the self-relevant condition.

Method

We adopted a 2 (Behavior domain: morality vs. ability) × 2 (Relevance of target individuals: self-relevant vs. self-irrelevant) × 2 (Behavior extremeness: highly negative vs. mildly negative) mixed design, with the latter factor manipulated within participants. The main dependent variable was generalization to judgments made on the alternative dimension (i.e., the Horn effect). We also measured to what extent each participant felt threatened by the four target individuals.

Sensitivity Analysis. We conducted a sensitivity analysis using a Monte Carlo simulation approach with the *Superpower* package in R (Lakens & Caldwell, 2021). A total of 1,000 simulations were run with a significance level set at 0.05. The simulations were based on the 2 × 2 × 2 mixed design and hypotheses described above, and we used the ANOVA_power() function to estimate the power to detect the main effects and interactions. An overall sample of 703 valid-data participants provided effectively 100% power to find an effect as small as $\eta^2 = .025$ for the main effect of behavior domain (i.e., the asymmetric Horn effect), effectively 100% power to find an effect as small as $\eta^2 = .015$ for the interaction between the Horn asymmetry and extremeness, and 76% power to find an effect as small as $\eta^2 = .003$ for the interaction between the Horn asymmetry and self-relevance, and 87% of power to find an effect as small as $\eta^2 = .003$ for the three-way interaction. The estimates differ from those preregistered, as applying the exclusion criteria resulted in a final sample smaller than planned.

Participants and Procedure. Eight hundred and six participants (382 females, 409 males, 9 unspecified, M_{age} = 41.09 years old, SD_{age} = 12.76 years old) entered the study via Prolific Academic. After providing informed consent, participants were randomly assigned to one of the four conditions derived from crossing behavior domain and relevance of the target individual. In the first part of the experiment, participants saw a list of 45 behaviors, belonging to either the immorality or the inability domain. Participants evaluated each behavior on the relevant dimension on an 11-point scale (−5 = *extremely negative*; +5 = *extremely positive*). For each participant, we retained two behaviors rated as −4 or −5 (i.e., highly negative) and two behaviors rated as −2 or −1 (i.e., mildly negative) for the following phase. We excluded participants who did not assign the critical scores to at least one behavior ($N = 103$).

Next, participants saw four individuals, each associated with one of the four behaviors (i.e., two mildly negative ones and two highly negative ones). Using a 11-point scales (ranging from −5 to +5), participants rated each individual on the alternative dimension (e.g., ability if the associated behaviors were immoral) and reported their level of threat (this scale ranged from −5 = *not at all* to 5 = *extremely*). In the self-relevant condition, we induced participants to believe that the four target individuals were people they would interact with later in the study. In the other self-irrelevant condition, we presented target individuals as random strangers.

Results

Generalization. We conducted a 2 (Behavior domain: morality vs. ability) × 2 (Relevance of target individual: self-relevant vs. self-irrelevant) × 2 (Behavior extremeness: highly negative vs. mildly negative) mixed ANOVA, with judgments on the alternative dimension as the DV. The effect of behavior domain was significant, $F(1,699) = 68.12$, $p < .001$, $\eta^2 = .065$. Target individuals associated with immoral behaviors were judged more negatively on ability ($M = -1.51$, $SD = 1.78$) than targets associated with unable behaviors were judged on morality ($M = -0.55$, $SD = 1.84$). There

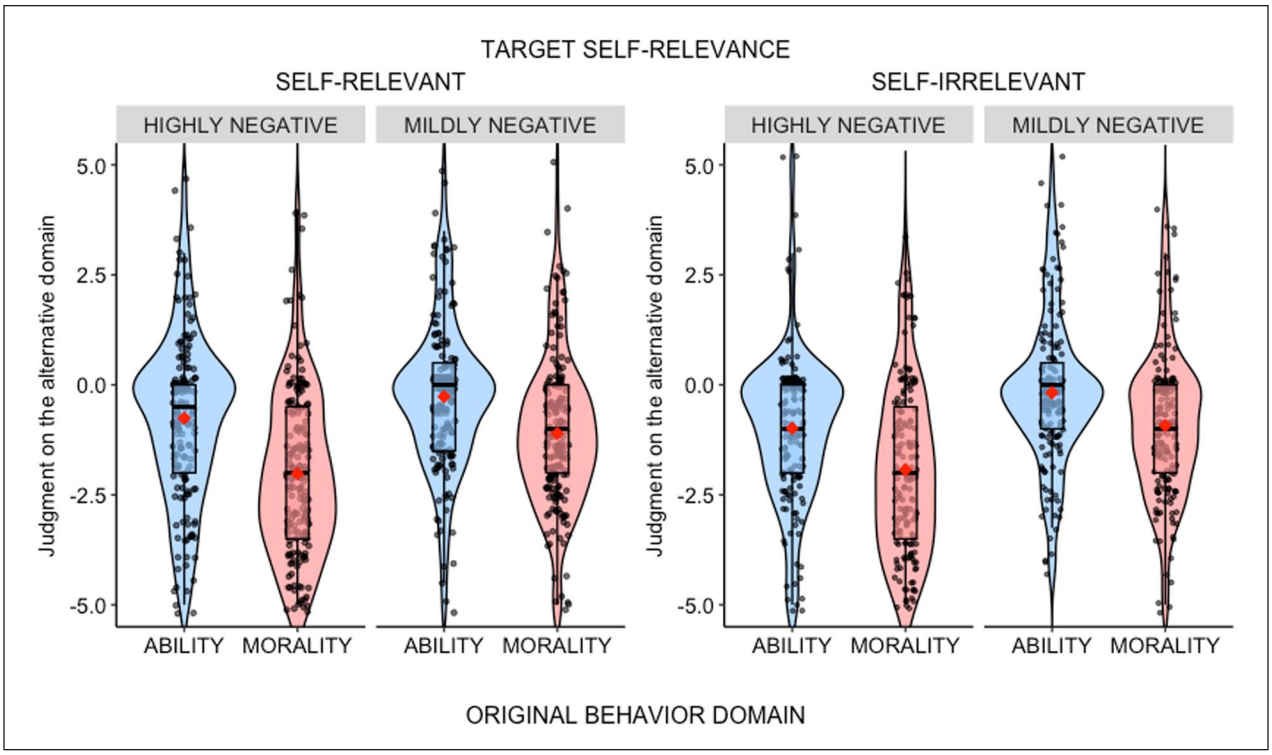

Figure 4. Effect of original behavior domain on between-dimension generalization for highly negative and mildly negative behaviors, by target self-relevance—Experiment 3.

Note. Boxplots are the interquartile range, and the bars represent the median; the red diamonds represent the mean. The distributions are the kernel probability density of the data.

was a significant effect of behavior extremeness, $F(1,699) = 152.20$, $p < .001$, $\eta^2 = .047$, indicating that highly negative behaviors produced stronger generalization ($M = -1.44$, $SD = 1.91$) than mildly negative ones ($M = -0.64$, $SD = 1.75$). No main effect of target personal relevance emerged, $F(1,699) = 0.09$, $p = .766$, $\eta^2 < .001$. The interaction between behavior domain and extremeness was significant, $F(1,699) = 5.78$, $p = .017$, $\eta^2 = .002$. Decomposing this interaction revealed that the asymmetric Horn effect (i.e., higher generalization from immoral behaviors to unable judgments) was stronger for highly negative behaviors, $F(1,701) = 64.85$, $p < .001$, $\eta^2 = .085$, than for mildly negative ones, $F(1,701) = 39.11$, $p < .001$, $\eta^2 = .053$. No other interactions reached significance (Fs < 2.41, ps $> .121$) (see Figure 4).

Perceived Threat. We repeated the same analyses on perceived threat. The effect of behavior domain was significant, $F(1,699) = 152.69$, $p < .001$, $\eta^2 = .145$, indicating that participants felt more threatened by targets associated with immoral behaviors ($M = -0.90$, $SD = 2.56$) than targets associated with unable behaviors ($M = -2.83$, $SD = 2.13$). The effect of behavior extremeness was also significant, $F(1,699) = 73.60$, $p < .001$, $\eta^2 = .017$, with targets associated with highly negative behaviors inducing higher threat ($M = -1.51$, $SD = 2.70$) than those associated with mildly negative behaviors ($M = -2.18$, $SD = 2.35$). We found no main effect of relevance of the target individual, $F(1,699) = 0.77$, $p = .380$, $\eta^2 = .001$. The interaction between behavior domain and extremeness was significant, $F(1,699) = 39.09$, $p < .001$, $\eta^2 = .009$. Decomposing this interaction revealed that the difference in threat generated by immoral as opposed to unable behaviors was stronger for the highly negative behaviors, $F(1,701) = 177.97$, $p < .001$, $\eta^2 = .202$, compared to the mildly negative behaviors, $F(1,701) = 75.24$, $p < .001$, $\eta^2 = .097$. No other interaction reached significance (see Figure 5).

Mediation Analysis. To investigate whether perceived threat statistically mediated the immorality (vs. inability) dominance in the Horn effect, we conducted a stepwise linear mixed-effects regression analysis using the *lmer* package in R (Bates et al., 2015).

We first tested the total effect of behavior domain on Horn generalization (path c) with a 3-way interaction model. Behavior domain significantly predicted generalization, $b = -0.95$, $F(1,703.68) = 67.34$, $p < .001$, indicating that immoral behaviors produced stronger generalization onto ability ratings than unable behaviors did onto morality ratings. The other significant effects were the main effect of

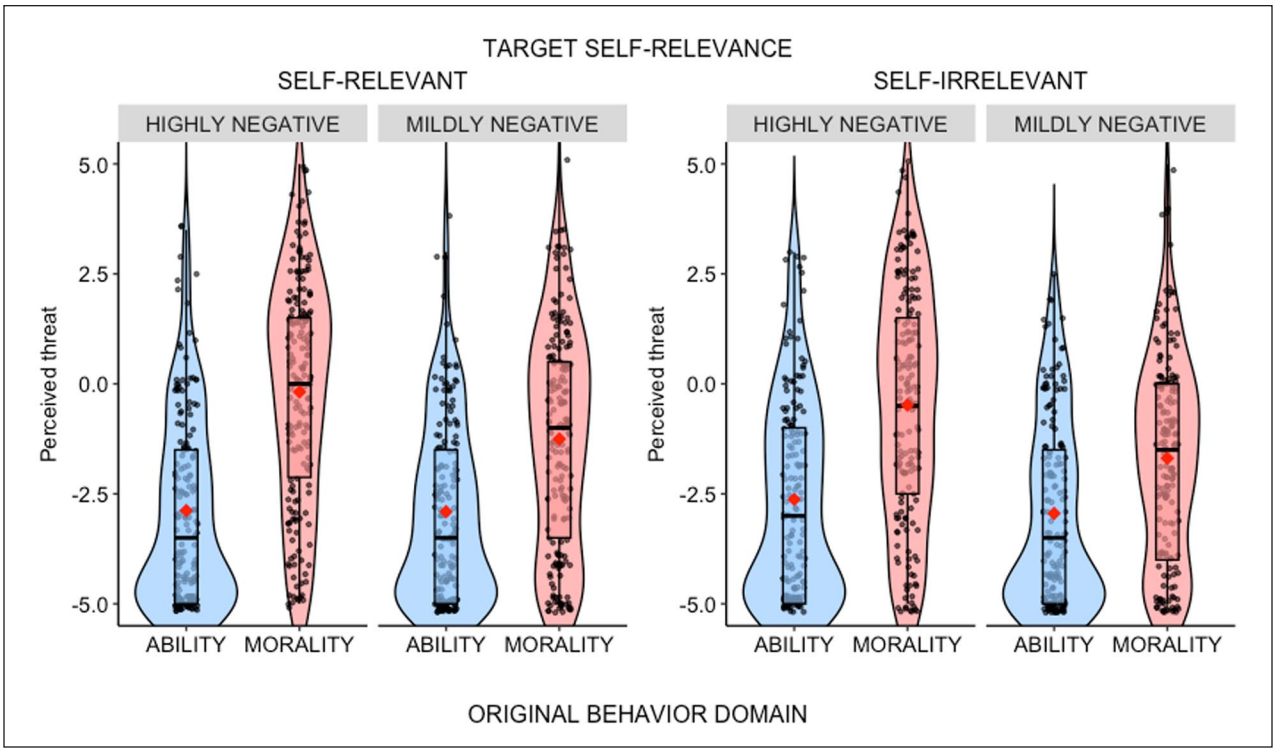

Figure 5. Effect of original behavior domain on perceived threat for highly negative and mildly negative behaviors, by target self-relevance—Experiment 3.

Note. Boxplots are the interquartile range, and the bars represent the median; the red diamonds represent the mean. The distributions are the kernel probability density of the data.

extremeness, $b = -0.77$, $F(1,720.39) = 145.63$, $p < .001$, and the interaction between behavior domain and extremeness, $b = -0.23$, $F(1,720.39) = 6.24$, $p = .013$.

Next, we tested whether behavior domain predicted perceived threat (path a), which it did, $b = 2.13$, $F(1,697.81) = 151.63$, $p < .001$. There was also a significant main effect of extremeness, $b = 0.33$, $F(1,714.21) = 70.51$, $p < .001$, and an interaction between behavior domain and extremeness, $b = 0.85$, $F(1,714.21) = 36.33$, $p < .001$.

Finally, we added perceived threat as a covariate to test its possible role in mediating the effect of behavior domain on generalization. Perceived threat significantly predicted Horn generalization (path b), $b = -0.18$, $F(1,1359.30) = 48.87$, $p < .001$. Higher perceived threat was associated with stronger Horn generalization. The re-estimated effect of behavior domain on Horn generalization was reduced but remained significant when perceived threat was included, $b = -0.54$, $F(1,950.34) = 13.03$, $p < .001$.

To further test the statistical mediation, we conducted a Monte Carlo simulation with 10,000 iterations to estimate the indirect effect of behavior domain (immorality vs. inability) on Horn generalization through perceived threat. The coefficients and standard errors for the a-path (effect of behavior domain on perceived threat) and b-path (effect of perceived threat on Horn generalization) were extracted from the models (see Figure 6). The indirect effect was significant, $b = -0.38$ [−0.56, −0.23] (see Figure 6).

Discussion

Experiment 3 aimed at conducting a rigorous test of the asymmetric Horn effect and clarify the role of behavior extremeness and relevance of the target individual as potential moderators and mechanisms underlying this asymmetry. The results provided strong evidence for the hypothesized asymmetry: immoral behaviors elicited significantly stronger generalization to person-level judgments of ability than unable behaviors did to person-level judgments of morality. This asymmetric Horn effect was larger for highly negative behaviors compared to mildly negative ones. That is, the extremeness of the negative behaviors amplified the Horn effect. However, no effect of the self-relevance emerged: immoral (vs. unable) behaviors produced a greater Horn effect regardless of whether the target was presented as someone relevant or irrelevant for the self. Because we did not check the effectiveness of our self-relevance manipulation, the null effect of self-relevance does not necessarily speak against the affective-motivational self-interest account that predicts the asymmetric Horn effect that we observed. It could be, for example, that the participants did not believe

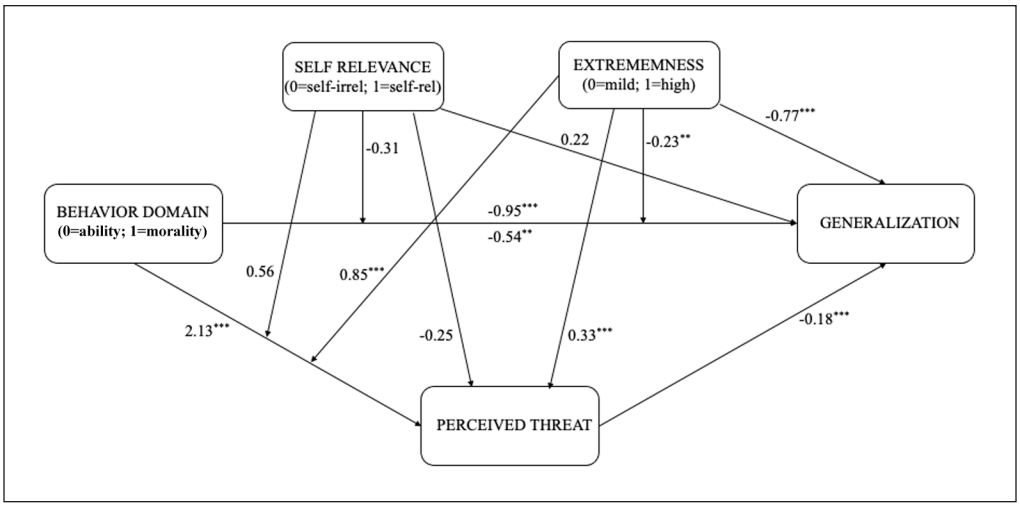

Figure 6. Mediation analysis—Experiment 3.

that they would actually interact with the target individuals in the self-relevant condition, or they believed so but anticipated a brief and mostly non-consequential, self-irrelevant social interaction.

Moreover, we found that targets associated with immoral behaviors were perceived as more threatening than those associated with inability. A mediation analysis further revealed that perceived threat partially accounted for the effect of behavior domain (morality vs. ability) on Horn generalization.

General Discussion

We investigated the mutual influence between morality and ability in social judgment, examining how behavioral information about one dimension affects inferences about the other. In Experiment 1, we studied generalization at the behavior level (i.e., participants self-generated behaviors that were positive and negative in one domain and judged those self-generated behaviors on the alternative domain). The results showed an asymmetry in cross-dimensional generalization only for the negative behaviors: immoral behaviors had a stronger impact on ability judgments than inability behaviors had on moral judgments (i.e., we found an asymmetric Horn effect). No Halo asymmetry emerged on the positive end. Experiments 2a and 2b extended the asymmetric Horn effect from the level of behaviors to persons (i.e., people generalized from behaviors to persons). Additionally, Study 2b provided insights into within-domain generalization, revealing that extremely immoral behaviors produced greater inferences of low sociability, compared to extremely unable behaviors producing inferences of low assertiveness. Highly powered Experiment 3 replicated the person-level asymmetric Horn effect and showed that the extremeness of the immoral (vs. unable) behaviors amplified the asymmetry. The largest asymmetry effect emerged when comparing the generalization starting form highly immoral versus highly unable behaviors. Moreover, we delved into the affective-motivational forces that might account for the asymmetry in the Horn effect. We found that immoral behaviors were seen as more threatening than unable behaviors, which partially statistically mediated people's greater tendency to infer inability from immoral behaviors (vs. inferring immorality from unable behaviors).

These findings make a novel contribution to the literature on the implications of perceived morality for impression formation. Most prior studies in this area have documented a relatively higher importance of morality (vs. other dimensions) by comparing the dimensions' impact on the global evaluation (i.e., impression) of other people (Brambilla et al., 2021; Goodwin, 2015). Our work extends this perspective by revealing that morality's influence is not only confined to overall impressions but also emerges in cross-dimensional inferences. At least on the negative side, immorality more readily spills over into ability judgments than ability spills over into morality judgments. We tentatively propose that this asymmetry in cross-dimensional generalization may help explain morality's broader influence in overall impression formation. Specifically, if both morality and ability are central to impression formation, the stronger tendency of negative moral information to elicit inferences in other core evaluative domains may indirectly contribute to the primacy of morality in global evaluations. Although our studies did not include a direct measure of overall impressions and the interpretation remains speculative, it is supported indirectly by the pattern observed in Study 2b: highly immoral behaviors yielded a stronger generalization to low sociability than highly unable behaviors did to low assertiveness. This asymmetry suggests that the impact of immoral behavior on other domains is more potent than that of inability on its corresponding domains. We therefore argue that highly immoral behavior may disproportionately shape overall evaluations,

and that cross-dimensional generalization—particularly for negative moral information—may serve as a pathway through which morality exerts broader influence in social judgment. Future studies could test this possibility in multiple ways: correlationally, by examining whether an asymmetry in generalization is associated with a greater weight for morality in influencing global impressions, and experimentally, by testing whether preventing asymmetric generalization reduces the superior role of morality in shaping overall evaluations.

Our contribution highlights the role of valence in how people generalize impressions. Consistent with prior research (Gräf & Unkelbach, 2016), in Experiment 1 (and 1b), we found that people generalized more strongly from positive than from negative information. This finding is consistent with the Evaluative Information Ecology (EvIE) Model (Unkelbach et al., 2019), which suggests that negative things are more diverse and less similar to one another than positive things, making it easier for people to generalize from positive impressions. Importantly, we also found an interaction between valence and behavior domain: asymmetry in cross-dimensional generalization occurred only for negative behaviors (an asymmetric Horn but not Halo effect). This result that not all negative information has the same impact is not predicted by the EvIE model. But prior work supports this pattern, showing that negative morality tends to exert a greater influence than negative morality-irrelevant information in shaping first impressions, behavioral responses, and impression updating (Brambilla et al., 2016; Luttrell et al., 2022; Menegatti et al., 2020; Reeder & Coovert, 1986; see also Brambilla et al., 2021). Moreover, this differential asymmetry in generalization contrasts with another cognitive ecological account of social judgments—the rarity account (Mende-Siedlecki et al., 2013; Skowronski & Carlston, 1987; Trafimow & Trafimow, 1999). This account predicts that because ability and immorality are less frequent than morality and inability, people should generalize more from ability to morality than vice versa. Yet, in our studies able behaviors did not produce greater impact on judgment of morality as opposed to what a moral behavior did on judgments of ability. This aligns with prior work suggesting that ability remains a dimension with relatively little impact on overall impressions (Brambilla et al., 2021; Rusconi et al., 2020) Together, these findings suggest that while cognitive-ecological models like EvIE and the rarity account explain part of the asymmetry, affective-motivational processes, such as heightened sensitivity to moral threats, also play a key role (see also Alves et al., 2017).

Our findings also make an original contribution to the literature on the reciprocal influence between morality and ability. Prior work has shown that negative moral behaviors generalize to negative judgments on the ability dimension (Stellar & Willer, 2018), but the boundary conditions of this effect remained obscure. In the current set of studies, we unveiled the role of (a) valence (i.e., positive vs. negative behaviors), (b) original behavior domain (i.e., from [im] moral behaviors to ability judgments vs. from [in]ability behavior to morality judgments), and (c) extremeness (i.e., mildly extreme vs. highly extreme behaviors) on the tendency to use information that is pertinent for one domain and make inference on the other domain. That people generalize more when the original behavior is positive, as opposed to negative (regardless of behavior domain), is consistent with prior related research by Gräf and Unkelbach (2016). On the negative side, the role of initial behavior domain was key: participants generalized more from immoral behaviors to inability judgment than vice-versa.

Finally, both the Horn effect and its asymmetry in the negative end were qualified by extremeness. The Horn effect was larger for highly (vs. mildly) negative behaviors, and the Horn effect asymmetry was larger among highly (vs. mildly) negative behaviors. Prior work has documented the relevance of evaluative extremeness. For instance, extreme moral behaviors are weighted more than moderate moral behaviors when forming overall impressions. Moreover, trait inferences from extremely (vs. moderately) immoral behaviors are more resistant to change following contradictory evidence (for a review, Rusconi et al., 2020). Our work further showed that evaluative extremeness also plays a role in shaping Horn generalizations between dimensions.

As they stand, our findings provide preliminary insights into the mechanisms driving the asymmetry in cross-dimensional generalizations for negative behaviors. Partly supporting an affective-motivational account, we found that immoral behaviors elicited stronger perceptions of threat than inability behaviors, and this was associated with the larger Horn effect observed when generalizing from immorality (vs. inability). This pattern is consistent with the possibility that perceived threat contributes to the greater tendency to downgrade immoral targets on the ability dimension. From a motivational perspective, attributing inability to immoral actors may function as a protective strategy, aimed at reducing the potential harm posed by malevolent intentions by denying the capacity to enact them (see also Landy et al., 2016; Wojciszke et al., 1998). However, these results do not provide definitive support for a threat-based mechanism. Mediation analyses alone cannot establish underlying psychological processes (Fiedler et al., 2011), and the absence of a moderation effect of self-relevance in Experiment 3 further suggests that threat-related motives may not be the sole, or even primary, driver of the observed asymmetry (but see Experiment 3 Discussion for an important methodological caveat regarding the self-relevance manipulation).

Alternative explanations remain plausible. For instance, according to a moral-telos perspective on social identity (Christy & Schlegel, 2024), immoral actors are perceived as deviating from the fundamental moral purposes that human beings are generally expected to pursue. If immoral actors are seen as misaligned with these essential human aims, then inferring inability may serve a self-protective function not

only by alleviating personal threat, but also legitimizing avoidance or diminishing their relevance as potential interaction partners. Another possibility, more cognitive-conceptual in nature, is that morality itself may be viewed as a form of ability. Prior work on cross-domain generalization has suggested that moral behavior might signal broader interpersonal abilities or social intelligence (Stellar & Willer, 2018), such that immoral actors may be judged as lacking ability by virtue of these conceptual links. A possibility is that immorality was seen as reflecting inability in anticipating longer-term consequences—namely, the reputational or material costs that can outweigh short-term benefits. Taken together, our studies highlight perceived threat as a promising, but not exclusive, pathway through which immorality exerts disproportionate influence on trait generalization. Future research could build on the present findings by (a) adopting more sensitive manipulations of motivational states, or more ecologically valid inductions of self-relevance, to better test the role of threat, and (b) directly comparing multiple candidate mediators within the same design, thereby clarifying their relative explanatory power.

Conclusion

In sum, our research explored the differential impact of perceived morality and ability in influencing each other. Our results confirmed the primacy of morality in generalizations between social-evaluative dimensions. While positive behaviors tend to produce Halo-type generalization of equal size in both directions, immoral behaviors influenced judgments of ability more than unable behaviors influenced judgment of ability. This asymmetric Horn effect was amplified among highly (vs. mildly) negative behaviors, and it may be partially explainable through the larger perceived threat induced by individuals who were associated with immoral (vs. unable) behaviors.

ORCID iDs

Simone Mattavelli https://orcid.org/0000-0002-8934-8016
Marco Brambilla https://orcid.org/0000-0002-4774-3309
Alex Koch https://orcid.org/0000-0002-6267-8066
Marcos Dono https://orcid.org/0000-0002-6895-2433

Ethical Considerations

The experiments received formal approval by the internal Ethics Committee.

Consent to Participate

All participants gave their consent to participate to the studies.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement

All data, analysis code and materials for the four experiments are available on Open Science Framework (https://osf.io/rfa76/)

Supplemental Material

Supplemental material is available online with this article.

Notes

1. Our preregistrations (except for Study 3) did not follow the OSF or As Predicted templates. Instead, we uploaded detailed, time-stamped documents to OSF that fully specified our hypotheses, methods, and analysis plans. Although this format is less standardized than template-based formats, it nonetheless provides verifiable timestamps and a clear temporal record of our intentions prior to data collection. For clarity, we also provide a Supplementary table (Table S1) documenting any inconsistencies between study names in the uploaded documents and on OSF.
2. Facial stimuli codes: CFD-WM-014-002-N, CFD-WM-025-002-N, CFD-WM-028-003-N, CFD-WM-210-057-N.
3. Participants also rated each selected behavior on diagnosticity on a 7-point scale in Experiment 2a (1 = *not at all diagnostic*; 7 = *extremely diagnostic*) and on a 11-point scale in Experiment 2b (−5 = *not at all diagnostic*; +5 = *extremely diagnostic*). For the sake of brevity and simplicity, we report the analysis on this variable in the Supplemental Material.

References

Abele, A. E. (2003). The dynamics of masculine-agentic and feminine-communal traits: Findings from a prospective study. *Journal of Personality and Social Psychology*, *85*(4), 768–776. https://doi.org/10.1037/0022-3514.85.4.768

Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology*, *93*(5), 751–763. https://doi.org/10.1037/0022-3514.93.5.751

Abele, A. E., Ellemers, N., Fiske, S. T., Koch, A., & Yzerbyt, V. (2021). Navigating the social world: Toward an integrated framework for evaluating self, individuals, and groups. *Psychological Review*, *128*(2), 290–314. https://psycnet.apa.org/doi/10.1037/rev0000262

Abele, A. E., Hauke, N., Peters, K., Louvet, E., Szymkow, A., & Duan, Y. (2016). Facets of the fundamental content dimensions: Agency with competence and assertiveness—Communion with warmth and morality. *Frontiers in Psychology*, *7*, 1810. https://doi.org/10.3389/fpsyg.2016.01810

Alves, H., Koch, A., & Unkelbach, C. (2017). Why good is more alike than bad: Processing implications. *Trends in Cognitive Sciences*, *21*(2), 69–79. http://dx.doi.org/10.1016/j.tics.2016.12.006

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Brambilla, M., & Leach, C. W. (2014). On the importance of being moral: The distinctive role of morality in social judgment. *Social Cognition*, *32*(4), 397–408. https://psycnet.apa.org/doi/10.1521/soco.2014.32.4.397

Brambilla, M., Carraro, L., Castelli, L., & Sacchi, S. (2019). Changing impressions: Moral character dominates impression updating. *Journal of Experimental Social Psychology*, *82*, 64–73. https://psycnet.apa.org/doi/10.1016/j.jesp.2019.01.003

Brambilla, M., Sacchi, S., Menegatti, M., & Moscatelli, S. (2016). Honesty and dishonesty don't move together: Trait content information influences behavioral synchrony. *Journal of Nonverbal Behavior*, *40*, 171–186. https://psycnet.apa.org/doi/10.1007/s10919-016-0229-9

Brambilla, M., Sacchi, S., Pagliaro, S., & Ellemers, N. (2013). Morality and intergroup relations: Threats to safety and group image predict the desire to interact with outgroup and ingroup members. *Journal of Experimental Social Psychology*, *49*(5), 811–821. https://doi.org/10.1016/j.jesp.2013.04.005

Brambilla, M., Sacchi, S., Rusconi, P., & Goodwin, G. P. (2021). The primacy of morality in impression development: Theory, research, and future directions. *Advances in Experimental Social Psychology*, *64*, 187–262. https://doi.org/10.1016/bs.aesp.2021.03.001

Brambilla, M., Sacchi, S., Rusconi, P., Cherubini, P., & Yzerbyt, V. Y. (2012). You want to give a good impression? Be honest! Moral traits dominate group impression formation. *British Journal of Social Psychology*, *51*(1), 149–166. https://doi.org/10.1111/j.2044-8309.2010.02011.x

Cambon, L., Yzerbyt, V., & Yakimova, S. (2015). Compensation in intergroup relations: An investigation of its structural and strategic foundations. *British Journal of Social Psychology*, *54*(1), 140–158. https://doi.org/10.1111/bjso.12067

Carrier, A., Dompnier, B., & Yzerbyt, V. (2019). Of nice and mean: The personal relevance of others' competence drives perceptions of warmth. *Personality and Social Psychology Bulletin*, *45*(11), 1549–1562. https://psycnet.apa.org/doi/10.1177/0146167219835213

Chawinwit, N., & Koch, A. (2025). Ability, Assertiveness, Morality, and Friendliness? A Retrospective Validation of the Facet Model of Social Evaluation. *Social Psychological and Personality Science*, *0*(0). https://doi.org/10.1177/19485506251371342

Chen, F. F., Jing, Y., & Lee, J. M. (2014). The looks of a leader: Competent and trustworthy, but not dominant. *Journal of Experimental Social Psychology*, *51*, 27–33. https://doi.org/10.1016/j.jesp.2013.10.008

Christy, A. G., Kim, J., Vess, M., Schlegel, R. J., & Hicks, J. A. (2017). The reciprocal relationship between perceptions of moral goodness and knowledge of others' true selves. *Social Psychological and Personality Science*, *8*(8), 910–917. https://doi.org/10.1177/1948550617693061

Christy, A. G., & Schlegel, R. J. (2024). Perceiving persons and their purposes: Teleology, normativity, and personal identity. *Self and Identity*, *23*(1–2), 23–48. https://doi.org/10.1080/15298868.2024.2314921

Cook, C. L., Krems, J. A., & Kenrick, D. T. (2021). Fundamental motives illuminate a broad range of individual and cultural variations in thought and behavior. *Current Directions in Psychological Science*, *30*(3), 242–250. https://doi.org/10.1177/0963721421996690

De Bruin, E. N., & Van Lange, P. A. (1999). Impression formation and cooperative behavior. *European Journal of Social Psychology*, *29*(2–3), 305–328. https://psycnet.apa.org/doi/10.1002/(SICI)1099-0992(199903/05)29:2/3%3C305::AID-EJSP929%3E3.0.CO;2-R

De Bruin, E. N., & Van Lange, P. A. (2000). What people look for in others: Influences of the perceiver and the perceived on information selection. *Personality and Social Psychology Bulletin*, *26*(2), 206–219. https://doi.org/10.1177/0146167200264007

Durante, F., Tablante, C. B., & Fiske, S. T. (2017). Poor but warm, rich but cold (and competent): Social classes in the stereotype content model. *Journal of Social Issues*, *73*(1), 138–157. https://doi.org/10.1111/josi.12208

Ellemers, N. (2017). *Morality and the regulation of social behavior: Groups as moral anchors*. Routledge.

Ellemers, N., Pagliaro, S., & Barreto, M. (2013). Morality and behavioural regulation in groups: A social identity approach. *European Review of Social Psychology*, *24*(1), 160–193. https://doi.org/10.1080/10463283.2013.841490

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. https://doi.org/10.3758/BRM.41.4.1149

Fiedler, K., Schott, M., & Meiser, T. (2011). What mediation analysis can (not) do. *Journal of Experimental Social Psychology*, *47*(6), 1231–1236. https://doi.org/10.1016/j.jesp.2011.05.007

Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality and Social Psychology*, *38*(6), 889–906. https://psycnet.apa.org/doi/10.1037/0022-3514.38.6.889

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. https://psycnet.apa.org/doi/10.1037/0022-3514.82.6.878

Goodwin, G. P. (2015). Moral character in person perception. *Current Directions in Psychological Science*, *24*(1), 38–44. https://doi.org/10.1177/0963721414550709

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, *106*(1), 148–168. https://doi.org/10.1037/a0034726

Gräf, M., & Unkelbach, C. (2016). Halo effects in trait assessment depend on information valence: Why being honest makes you industrious, but lying does not make you lazy. *Personality and Social Psychology Bulletin*, *42*(3), 290–310. https://doi.org/10.1177/0146167215627137

Imhoff, R., & Koch, A. (2017). How orthogonal are the Big Two of social perception? On the curvilinear relation between agency and communion. *Perspectives on Psychological Science*, *12*(1), 122–137. https://psycnet.apa.org/doi/10.1177/1745691616657334

Judd, C. M., James-Hawkins, L., Yzerbyt, V., & Kashima, Y. (2005). Fundamental dimensions of social judgment: Understanding

the relations between judgments of competence and warmth. *Journal of Personality and Social Psychology*, *89*(6), 899-913. https://doi.org/10.1037/0022-3514.89.6.899

Kenrick, D. T., Neuberg, S. L., Griskevicius, V., Becker, D. V., & Schaller, M. (2010). Goal-driven cognition and functional behavior: The fundamental-motives framework. *Current Directions in Psychological Science*, *19*(1), 63–67. https://doi.org/10.1177/0963721409359281

Koch, A., Imhoff, R., Dotsch, R., Unkelbach, C., & Alves, H. (2016). The ABC of stereotypes about groups: Agency/socioeconomic success, conservative–progressive beliefs, and communion. *Journal of Personality and Social Psychology*, *110*(5), 675-709. https://doi.org/10.1037/pspa0000046

Koch, A., Imhoff, R., Unkelbach, C., Nicolas, G., Fiske, S., Terache, J., Carrier, A., & Yzerbyt, V. (2020). Groups' warmth is a personal matter: Understanding consensus on stereotype dimensions reconciles adversarial models of social evaluation. *Journal of Experimental Social Psychology*, *89*, 103995. https://doi.org/10.1016/j.jesp.2020.103995

Koch, A., Smith, A., Fiske, S. T., Abele, A. E., Ellemers, N., & Yzerbyt, V. (2024). Validating a brief measure of four facets of social evaluation. *Behavior Research Methods*, *56*(8), 8521–8539. https://doi.org/10.3758/s13428-024-02489-y

Koch, A., Yzerbyt, V., Abele, A., Ellemers, N., & Fiske, S. T. (2021). Social evaluation: Comparing models across interpersonal, intragroup, intergroup, several-group, and many-group contexts. *Advances in Experimental Social Psychology*, *63*, 1-68. https://doi.org/10.1016/bs.aesp.2020.11.001

Lakens, D., & Caldwell, A. R. (2021). Simulation-based power analysis for factorial analysis of variance designs. *Advances in Methods and Practices in Psychological Science*, *4*(1). https://doi.org/10.1177/2515245920951503

Landy, J. F., Piazza, J., & Goodwin, G. P. (2016). When it's bad to be friendly and smart: The desirability of sociability and competence depends on morality. *Personality and Social Psychology Bulletin*, *42*(9), 1272–1290. https://doi.org/10.1177/0146167216655984

Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: The importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of Personality and Social Psychology*, *93*(2), 234–249. https://doi.org/10.1037/0022-3514.93.2.234

Luttrell, A., Sacchi, S., & Brambilla, M. (2022). Changing impressions in competence-oriented domains: The primacy of morality endures. *Journal of Experimental Social Psychology*, *98*, 104246. https://doi.org/10.1016/j.jesp.2021.104246

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135.

Mende-Siedlecki, P., Said, C. P., & Todorov, A. (2013). The social evaluation of faces: A meta-analysis of functional neuroimaging studies. *Social Cognitive and Affective Neuroscience*, *8*(3), 285–299. https://doi.org/10.1093/scan/nsr090

Menegatti, M., Moscatelli, S., Brambilla, M., & Sacchi, S. (2020). The honest mirror: Morality as a moderator of spontaneous behavioral mimicry. *European Journal of Social Psychology*, *50*(7), 1394–1405. https://doi.org/10.1002/ejsp.2670

Nicolas, G., Bai, X., & Fiske, S. T. (2022). A spontaneous stereotype content model: Taxonomy, properties, and prediction. *Journal of Personality and Social Psychology*, *123*(6), 1243–1263. https://psycnet.apa.org/doi/10.1037/pspa0000312

Pagliaro, S., Ellemers, N., & Barreto, M. (2011). Sharing moral values: Anticipated ingroup respect as a determinant of adherence to morality-based (but not competence-based) group norms. *Personality and Social Psychology Bulletin*, *37*(8), 1117–1129. https://doi.org/10.1177/0146167211406906

Reeder, G. D., & Brewer, M. B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review*, *86*(1), 61–79. https://psycnet.apa.org/doi/10.1037/0033-295X.86.1.61

Reeder, G. D., & Coovert, M. D. (1986). Revising an impression of morality. *Social Cognition*, *4*(1), 1–17. https://doi.org/10.1521/soco.1986.4.1.1

Roberts, R., & Koch, A. (2025). Perceived power polarizes moral evaluations. *Personality and Social Psychology Bulletin*, *51*(11), 2150–2165. https://doi.org/10.1177/01461672241245181

Rusconi, P., Sacchi, S., Brambilla, M., Capellini, R., & Cherubini, P. (2020). Being honest and acting consistently: Boundary conditions of the negativity effect in the attribution of morality. *Social Cognition*, *38*(2), 146–178. https://psycnet.apa.org/doi/10.1521/soco.2020.38.2.146

Skowronski, J. J., & Carlston, D. E. (1987). Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. *Journal of Personality and Social Psychology*, *52*(4), 689–699. https://psycnet.apa.org/doi/10.1037/0022-3514.52.4.689

Stellar, J. E., & Willer, R. (2018). Unethical and inept? The influence of moral information on perceptions of competence. *Journal of Personality and Social Psychology*, *114*(2), 195–210. https://psycnet.apa.org/doi/10.1037/pspa0000097

Stolier, R. M., Hehman, E., & Freeman, J. B. (2020). Trait knowledge forms a common structure across social cognition. *Nature Human Behaviour*, *4*(4), 361–371. https://doi.org/10.1038/s41562-019-0800-6

Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition*, *131*, 159–171. https://doi.org/10.1016/j.cognition.2013.12.005

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology 4*, 25–29. https://psycnet.apa.org/doi/10.1037/h0071663

Trafimow, D., & Trafimow, S. (1999). Mapping perfect and imperfect duties onto hierarchically and partially restrictive trait dimensions. *Personality and Social Psychology Bulletin*, *25*(6), 687–697. https://psycnet.apa.org/doi/10.1177/0146167299025006004

Unkelbach, C., Koch, A., & Alves, H. (2019). The evaluative information ecology: On the frequency and diversity of "good" and "bad." *European Review of Social Psychology*, *30*(1), 216–270. https://doi.org/10.1080/10463283.2019.1688474

Wojciszke, B. (2005). Morality and competence in person-and self-perception. *European Review of Social Psychology*, *16*(1), 155–188. https://doi.org/10.1080/10463280500229619

Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Personality and Social Psychology Bulletin*, *24*(12), 1251–1263. https://doi.org/10.1177/01461672982412001

Wojciszke, B., Baryla, W., Parzuchowski, M., Szymkow, A., & Abele, A. E. (2011). Self-esteem is dominated by agentic

over communal information. *European Journal of Social Psychology*, *41*(5), 617–627. https://doi.org/10.1002/ejsp.791

Yzerbyt, V. (2016). Intergroup stereotyping. *Current Opinion in Psychology*, *11*, 90–95. https://doi.org/10.1016/j.copsyc.2016.06.009

Yzerbyt, V. (2018). The dimensional compensation model: Reality and strategic constraints on warmth and competence in intergroup perceptions. In *Agency and Communion in Social Psychology* (pp. 126–141). Routledge.

Yzerbyt, V., & Cambon, L. (2017). The dynamics of compensation: When ingroup favoritism paves the way for outgroup praise. *Personality and Social Psychology Bulletin*, *43*(5), 587–600. https://doi.org/10.1177/0146167216689066

Yzerbyt, V. Y., Kervyn, N., & Judd, C. M. (2008). Compensation versus halo: The unique relations between the fundamental dimensions of social judgment. *Personality and Social Psychology Bulletin*, *34*(8), 1110–1123. https://psycnet.apa.org/doi/10.1177/0146167208318602

Yzerbyt, V., Koch, A., Brambilla, M., Ellemers, N., Fiske, S. T., & Nicolas, G. (2025). Dimensions of stereotypes about groups. *Current Directions in Psychological Science*, *34*(6), 365–372. https://doi.org/10.1177/09637214251345489

Yzerbyt, V., Provost, V., & Corneille, O. (2005). Not competent but warm... really? Compensatory stereotypes in the French-speaking world. *Group Processes & Intergroup Relations*, *8*(3), 291–308. https://doi.org/10.1177/1368430205053944