Characterizing and Explaining Moral Perceptions of the Self, Individuals, and Collectives

WORD COUNT (Introductions + Discussions): 3,716

Authors' note: Materials, data, analysis code and preregistrations can be accessed online: https://osf.io/mg5yk/?view_only=bb8cdbfa0e2f403497adba9b8819c5d7.

Submission date: January 23rd, 2024

Abstract

For decades, psychologists have appreciated that the average person sees themselves as better than average. This is particularly true in moral domains. Although self-other comparisons are useful for establishing normative violations, they leave unanswered whether people see the self and others as positive and moral, or negative and immoral, in an absolute sense. The present research introduces a novel measure of *moral thresholds* to identify the behavioral tipping point that subjectively differentiates morality from immorality. In two different cultural contexts, the self was found to view itself as clearly moral while it viewed others (in the study) as falling short of the moral threshold, though less consistently so (Studies 1 and 2). Of course, social targets can take different forms. Study 3 found that although collectives (others in the study or in society) were seen to fall short of moral thresholds, individuals—even unknown specific others—were judged as exceeding moral thresholds. Studies 4a and 4b used a causal-chain design to explain why. People anticipated feeling worse from being cynical about an individual (as opposed to a collective). These anticipated negative feelings were then causally responsible for more positive behavioral forecasts. The moral threshold measure allows moral perception to join other domains (e.g., monetary outcomes, attitudes) in which identifying a neutral reference point has been core to future theoretical and empirical development. In addition to identifying newly addressable questions, discussion focuses on how the findings comport with certain previously established theories or phenomena while suggesting the need to revisit others.

Keywords: self-enhancement, better-than-average effect, moral perception, identifiable victim effect, error management theory

Statement of Limitations

We identify three primary limitations. First, each study was conducted in one of two cultural contexts, one of two countries located on different continents. Although we observed quite similar patterns of results, both are Western, industrialized nations. Replications in more cultural contexts would be necessary to identify potential cross-cultural variability. Second, each study's materials drew on a subset of 12 everyday moral and 12 everyday immoral behaviors. Some of these materials were selected through a multi-stage process designed to avoid experimenter bias in stimulus selection, but these behaviors may not be representative of the full array of all moral and immoral behaviors that could have been selected. Replication with even more behaviors would strengthen confidence in the present work's conclusions. Third, Studies 1 and 2 both yielded consistent support for self-positivity but less consistent (though robust in-the-aggregate for each study) support for other-negativity (when considering others in the study as a collective). There is thus ambiguity as to whether observed inconsistencies in when other-negativity emerged are explained by cultural differences between the two samples or instead simply random variability. Given other-negativity was smaller in magnitude than self-positivity, noise has a greater potential to introduce variability in when other-negativity emerges.

Characterizing and Explaining Moral Perceptions of the Self, Individuals, and Collectives

The average person sees themselves as better than average. People judge flattering personality traits to be more characteristic of the self and negative traits to be more characteristic of others (Brown, 1986). The self sees itself as more competent (Kruger & Dunning, 1999), more objective (Kruger & Gilovich, 1999), and less biased (Pronin et al., 2002) than others. These patterns have been observed across myriad contexts, domains, and methodologies (Zell et al., 2020).

Although self-enhancement extends to domains as diverse as one's intelligence, driving ability, tech-savviness, and emotional stability (Brown, 2012; Horswill et al., 2004; Kruger, 1999; Kruger et al., 2008; Zell & Alicke, 2011), it is particularly robust in moral domains (Tappin & McKay, 2016; Van Lange & Sedikides, 1998). In general, people expect others to be less likely to act morally (Epley & Dunning, 2000) and more capable of acting immorally than they do themselves (Klein & Epley, 2016, 2017).

This decades-long interest in the better-than-average effect, and people's sense of moral superiority in particular, has often overlooked that such phenomena blur two distinct evaluations. Self- and social perceptions combine to produce beliefs that the self is better than others. Researchers often solicit self-other comparisons directly (e.g., "Would you say you are more or less moral than others in this study?") because they facilitate identification of people as biased (cf. Critcher et al., 2011). But such comparisons give little insight into whether people see themselves and others as morally outstanding, inadequate, or neither.

The present work asks whether people—in seeing themselves as more likely than others to engage in good, moral behaviors and less likely to engage in bad, immoral behaviors—see themselves and others as moral (displaying what we call self- or other-positivity), immoral (displaying self- or other-negativity), or as morally neutral. In this way, we differentiate this work from previous research that asked whether people self-enhance or other-derogate—i.e., whether assessments and forecasts are inaccurately positive or pessimistic (Balcetis & Dunning, 2013; Epley & Dunning, 2000, 2006; Helzer & Dunning, 2012). Although studies of enhancement and derogation can identify when perceivers are likely to be surprised by their own or others' observed behavior, such methodologies are unable to identify whether people see their own or others' behavior as moral or immoral in an absolute sense. After all, one can underestimate how much a target will give to charity, but still believe that lowball estimate is sufficiently generous to merit moral praise. What has been missing is a measure of people's beliefs about the thresholds or tipping points that differentiate a good or acceptable level of a behavior from a subpar one. By introducing such a measure, this paper can identify whether perceptions of the self and others tend to reach or exceed such a moral threshold.

One complexity to this research question is that "others" can take many forms (e.g., others in general, a specific other, a known other). Thus, as part of addressing whether people display other-positivity or other-negativity, we explore how the nature of the other matters. In the process, we identify a novel psychological mechanism that explains why there is predictable variability in social perceptions tied to the nature of the target.

Contributors to Positivity and Negativity in Self and Social Judgment

Although previous research does not directly address whether the self and others are viewed as fundamentally moral or immoral, extensive research has identified psychological mechanisms that contribute directionally to the positivity or negativity of those assessments. For example, people disproportionately focus on their own personal contributions to joint endeavors (Kruger & Savitsky, 2009), which can elevate self-positivity when contributions are positive and plentiful, but temper positive assessments when contributions are negative and rare (Kruger & Gilovich, 1999; Kruger & Savitsky, 2009). People's recollections are dominated by their own frequent, positive behaviors and others' rarer, negative behaviors (Messick et al., 1985). And even when people's actions fall short of their personal standards, they fail to remember just how selfish they were (Carlson et al., 2020). When looking to the future, people base their own forecasts on their lofty (but often unrealized) positive intentions, whereas their social judgments hew closer to actually observed behavioral histories (Epley & Dunning, 2000, 2006; Helzer & Dunning, 2012; Kruger & Gilovich, 2004; Steimer & Mata, 2016; Williams & Gilovich, 2012). On balance, these forces push for positive self-judgments; this encourages, but does not guarantee, self-positivity. After all, the very existence of goal setting, aspirational ideal and ought selves (Higgins, 1987; Shah et al., 1998), and perceived prospects for improvement and change (Steimer & Mata, 2016) highlights how actual self-views may not necessarily be seen as reaching one's own standards.

Other research implicates forces that depress social perceptions. Perceivers see immorality in others' morally ambiguous actions (Hester et al., 2020) and engage in attributional cynicism to explain them (Critcher & Dunning, 2011). More generally, people buy into a norm that self-interest guides behavior (Miller, 1999; Miller & Ratner, 1998; Wenzel, 2005). This research highlights why people arrive at less morally charitable views of others.

All these literatures identify various mechanisms that have a directional influence on self and social perception, but without a comparison standard that allows for the classification of beliefs and forecasts as perceptions of moral adequacy or inadequacy, it remains unknown how the self and social targets are perceived. We thus introduce a measure of *moral threshold*—the behavioral base rate that a perceiver identifies for a specific behavior as the tipping point between moral adequacy and inadequacy. A moral threshold is not an objective fact; it is a perceiver's subjective standard. When person perceptions systematically exceed thresholds, they reflect positivity; when they systematically fall short of them, they reflect negativity. Addressing whether self-assessments reflect positivity or negativity is straightforward but, given the diversity of forms that *others* can take, distinguishing other-positivity and other-negativity is more nuanced.

Preemptively Managing the Discomfort of Being Cynical about Individuals

In understanding whether people hold generally positive or negative moral perceptions of others, it likely matters who this “other” is. In most studies of self-enhancement, the exact nature of the “other” is something of an afterthought. After all, when the self directly compares itself against others, such comparative judgments are more a function of self-views than they are other-views (Klar & Giladi, 1999). The other is included largely to serve as a normative comparison standard for self-views.

There are a few notable exceptions. First, once the self actually develops a relationship with an other, many of the processes that produce self-enhancement are extended to these others who form the extended self (Murray, 1999). Second, Alicke et al. (1995) showed that people compare themselves more favorably to collections of people or group averages than they do to specific individuals. Third, Critcher and Dunning (2013) showed that a randomly selected individual was forecast to behave more morally than were members of the group from which the individual was drawn. Notable in the second and third examples is that the greater positivity afforded to individuals emerged in the absence of any individuating information. That is, although actually learning individuating information can lead perceivers to shed (often negative) social-category-based group stereotypes (e.g., Lammers et al., 2021), targets’ status as *mere individuals* appears sufficient to lead them to be seen as better, more moral people (see also Sears, 1983).

This literature provides hints as to which social targets may be judged more or less positively—with individuals (especially individuated ones) on one end of the continuum and collectives on the opposite end. But left unanswered is how these targets will be assessed with reference to perceivers' moral thresholds. This may be important if generalized others and merely individuated individuals find themselves not merely judged differently, but actually on opposite sides of the neutral reference point. If so, this could suggest that people are generally soured on the moral character of others even as they approach individual members of that collective with trust in their moral adequacy (Dunning et al., 2014).

Still left unanswered would be *why* individuals would be offered charitable characterizations even as the collectives those individuals compose are not. Critcher and Dunning (2014) proposed three explanations for individual-collective asymmetries. First, people focus on individual-level features when forecasting individuals (e.g., what an individual's moral conscience would compel one to do) but group-level, social features when forecasting collectives' behavior (e.g., what social norms compel one to do). This can produce divergences in social forecasts (Critcher & Dunning, 2013). Second, Critcher and Dunning (2014) make a functional argument. Baseline trust and cooperation are core to social relationships, and those relationships occur between individuals. People need to approach individuals with a certain optimism and respect in order to even have a chance to learn if they are good, trustworthy people who are worthy of investment (Fetchenhauer & Dunning, 2010). Of course, functional arguments more highlight the practical usefulness of phenomena than they do explain them.

It is here that Critcher and Dunning's (2014) third proposal—one that, to our knowledge, has not been previously tested—may offer a key explanation. They argue that people likely find negative judgments of specific individuals to seem harsher and thus more aversive than negative judgments of collectives. Relatedly, individuals' actions have been shown to evoke stronger emotional responses than collectives' (Walker & Gilovich, 2021). Closer to the present proposal, the *identifiable victim effect* highlights that the misfortune faced by an innocent individual can be more distressing than the same fate faced by a collective (Kogut & Ritov, 2005a, 2005b; Small & Loewenstein, 2003; Small et al., 2007). We suggest that perceivers anticipate these properties and are thus especially reluctant to *victimize* an individual by offering what could be an inaccurately harsh assessment of them. Much as people find it easier to dehumanize groups than they do specific individuals (Golebiowska, 2001), we suggest that perceivers' anticipation that they will experience greater aversive reactions to being wrong about individuals, as opposed to collectives, preemptively encourages perceivers to try to avoid victimizing individuals (with cynicism) in the first place.

Overview of the Current Studies

We present five studies that test whether perceivers bestow positivity or negativity on the self and others, consider how these conclusions depend on the nature of the other, and probe a previously proposed but never-tested account of these effects. Studies 1 and 2 introduce the new moral threshold measure to explore the presence of self and other positivity and negativity in two different cultural contexts. Study 3 systematically examines which social targets (e.g., individuated vs. nonindividuated, individual vs. collective) are viewed positively or negatively. Studies 4a and 4b use an experimental causal chain design to test whether greater anticipated reactions to being cynical about individuals (as opposed to collectives) cause a reduction in cynicism toward individuals.

Study 1

Method

Participants and Design. One hundred sixteen undergraduate students (88.8% female, 7.88% male, 2.6% non-binary, 0.9% no response; $M_{\text{age}} = 21.58$, $SD_{\text{age}} = 6.15$) from a large European university took part in the study in exchange for course credit. Participants were randomly assigned to one of three *target* conditions: self, others, or threshold. A second factor, *morality*, was varied within-subjects.

Procedure. Participants learned that they would be considering a set of morally relevant behaviors. What judgment participants offered about each behavior was determined by their target condition. For two of the conditions—self and others—participants estimated what percentage of the time, in the situation described, the target engages in the behavior. Those in the *self* condition reported on their own behavior. Those in the *others* condition offered an estimate about the others in the study. The behaviors were presented in a fully randomized order.

Participants assigned to the *threshold* condition made qualitatively different judgments. These participants read a detailed passage that introduced the concept of a "moral threshold"—i.e., the percentage prevalence of a behavior that would reflect the dividing line between whether people are morally good or morally bad. That is, the identified threshold was said to be that point such that if moral behavior exceeded (or immoral behavior fell short of) the threshold, this would reflect that people were morally good. But if moral behavior fell short of (or immoral behavior exceeded) the threshold, this would reflect moral badness. Participants identified such a tipping point for each behavior. The instructions were presented twice, modified slightly based on whether participants would be considering moral or immoral behaviors. We counterbalanced the order in which participants considered the set of moral and the set of immoral behaviors, and then randomized the presentation of the behaviors within each set.

Materials. All participants considered the same set of 14 behaviors: 7 moral, 7 immoral. Each behavior took the same form. It identified a context (e.g., "when having trash and there are no trashcans nearby", "when getting too much change after paying at a store or café") and asked about a behavior that could occur in that context (e.g., the percentage of the time one throws the trash on the ground; the percentage of the time one corrects the cashier and returns the extra money).

Each behavior was adapted into a self, others, and threshold judgment. For example, an item that asked about how often one recycles read “In the last year, what percentage of your recyclables did you actually recycle?” (self judgment), “In the last year, what percentage of others’ recyclables did they actually recycle?” (others judgment; “others’” was specified to be others in the study), or instead asked participants to indicate the recycling rate that differentiates moral from immoral behavior (threshold judgment). One behavior applied to a future event, so it was adapted to ask about the likelihood that the self or others would engage in it. The English translations of the full set of behaviors are provided in Appendix A.

We conducted a pilot study to validate that the everyday behaviors we generated did indeed differ as expected in their perceived morality. We asked 107 CloudResearch-approved Americans recruited from Amazon’s Mechanical Turk to rate each of the 14 behaviors on a 7-point scale in terms of its perceived morality (1 = *Very immoral*, 4 = *As immoral as it is moral*, 7 = *Very moral*). Although we counterbalanced the directionality of the scale, we present results such that higher numbers reflect greater perceived morality. As expected, the 7 moral behaviors were seen as more moral (M = 6.13, SD = 0.86) than the 7 immoral behaviors (M = 2.55, SD = 0.73), paired t(106) = 27.39, p < .001, d = 2.65. In fact, each behavior significantly differed from the neutral midpoint in the expected direction, ts > 8.71, ps < .001.

Transparency and Openness. For this and all studies in this manuscript, we report all data exclusions (if any), all manipulations, and all measures. The data, analysis code, research materials, and (if relevant) the preregistrations (for Study 4a, Study 4b pretest, and Study 4b) are available on the Open Science Framework: https://osf.io/mg5yk/?view_only=bb8cdbfa0e2f403497adba9b8819c5d7.

Results

In order to test for evidence of self and other positivity or negativity, we used mixed models that allow us to compare behavioral estimates for the self and others against the moral threshold. Toward this end, we characterized the *morality* of each behavior (+1 = moral, -1 = immoral). *Target* (self, others, or threshold) was included as a categorical predictor, as was its interaction with morality. Finally, we included two random factors to account for the non-independence of responses: one for *participant* and one for the specific *behavior* being judged.

Suggesting that self and/or other judgments likely did depart from the moral threshold (and thus display evidence of positivity or negativity), we observed a significant Morality X Target interaction, $F(2, 1489.32) = 263.23$, $p < .001$ (see Table 1). We decomposed this omnibus interaction into the three 2 X 2 interactions. To begin, we found that participants thought that they behaved more morally than did others in the study, $t(1489.32) = 21.37$, $p < .001$. This replicates the better-than-average effect. Decomposing this interaction further, the self thought that it engaged in moral behaviors more frequently, $B = 22.01$, $SE = 2.38$, $t(264.45) = 9.26$, $p < .001$, and in immoral behaviors less frequently than did the others, $B = -38.31$, $SE = 2.38$, $t(264.45) = -16.12$, $p < .001$. But does this reflect self-positivity, other-negativity, both, or some other combination?

Table 1

Mean (SD) Estimates by Target and Morality of Behaviors (Study 1)

Morality	Self	Others	Threshold
Moral Behaviors	82.04_a (26.37)	60.06_b (24.36)	59.47_b (24.97)
Immoral Behaviors	21.46_a (23.87)	59.78_c (25.21)	48.17_b (26.82)
Morality Composite	60.58_a	0.28_c	11.30_b

Note. The morality composite reflects the moral behavior composite minus the immoral behavior composite. Comparing the threshold composite against the self and others composites offers tests of overall self-positivity and other-negativity. Means in the same row that do not share the same subscript differ at the $p < .05$ level.

To test for self-positivity, we examined the Morality X Target (self or threshold) interaction. This self-threshold gap was particularly robust, illustrating that the self thought its own behavior clearly exceeded the threshold for morality, *t*(1489.32) = 17.84, *p* < .001. Specifically, the self indicated that it performed moral behaviors more frequently than the threshold required, *B* = 22.57, *SE* = 2.33, *t*(263.17) = 9.70, *p* < .001, and immoral behaviors less frequently than the threshold would permit, *B* = -26.71, *SE* = 2.38, *t*(263.17) = -11.48, *p* < .001.

We then proceeded to test for other-negativity. We again observed a Morality X Target (others or threshold) interaction, *t*(1489.32) = 3.94, *p* < .001. This offered evidence of other-negativity, but here the effect was asymmetric. Although participants thought others' immoral behaviors were more frequent than the threshold would permit, *B* = 11.69, *SE* = 2.36, *t*(264.47) = 4.91, *p* < .001, they did not think that others' moral behaviors fell below the moral threshold, *B* = 0.57, *SE* = 2.36, *t*(264.47) = 0.24, *p* = .814. This offers initial evidence that other-negativity, though present, may be less robust than self-positivity.

Study 2

Study 2 extended on Study 1 in two ways. First, we sought to replicate the findings in a different cultural context and with a non-university sample. Second, note the moral and immoral behaviors in Study 1 differed not only in their morality but in their actual content. Although this subtle confound is common in better-than-average-effect studies, Study 2 solved it by manipulating the framing of each behavior to take either the original frame or its inverse (e.g., the percentage of the time that a target does *not* engage in a behavior). This turns the originally framed moral and immoral behaviors into inversely framed immoral and moral behaviors, respectively.

Method

Participants and Design. Two hundred eighteen Americans (52.3% female, 46.3% male, 0.9% non-binary, 0.5% agender; M_{age} = 30.35, SD_{age} = 11.48) recruited through Prolific took part in the study. Participants were randomly assigned to one of six conditions in a 3 (target: self, others, or threshold) X 2 (frame: original or inverse) full-factorial design. A third factor, morality, was measured within-subjects. Note that the frame factor is primarily a counterbalancing factor that unconfounds the content of the behavior from its morality.

Procedure. Like in Study 1, participants considered 14 behaviors: 7 moral behaviors and 7 immoral behaviors. Those in the *self* condition indicated what percentage of the time they engage in each behavior. Those in the *others* condition instead estimated what percentage of the time the other participants in the study engage in each behavior. Finally, those in the *threshold* condition identified the cutoff point for each behavioral base rate that would reflect morality as opposed to immorality.

Participants responded to the items using either the *original* frame (as used in Study 1) or the *inverse* frame. For the inverse frame, each moral behavior was actually the opposite of an original immoral behavior, whereas each immoral behavior was actually the opposite of an original moral behavior. For example, those in the original frame conditions still answered questions about—as two examples—the percentage of the time they washed their hands after using the restroom (moral behavior) and the percentage of the time they threw trash on the ground (immoral behavior). Those in the inverse frame instead indicated the percentage of the time they did *not* wash their hands after using the restroom (immoral behavior) and the percentage of the time they hold onto trash until they can dispose of it instead of throwing it on the ground (moral behavior).

Results

We again tested for evidence of self-positivity and other-negativity, this time using a design in which the behavioral contexts were unconfounded from their morality. We used a mixed model that predicted participants' judgments. This model included a fixed effect of *morality* that indicated whether the behavior being judged was moral (+1) or immoral (-1). A fixed effect of *frame* differentiated those participants who considered the original (+1) set of behaviors (used in Study 1) from those who saw the inverse (-1) frame—i.e., the one in which the behavioral descriptions in the original frame were flipped so that the moral behaviors became immoral ones (and vice versa). In addition, *target* (self, others, threshold) was included as a categorical variable. The full set of interaction terms that could be created from these fixed effects was added. Finally, the model included two random effects to account for non-independence in the data: one for *participant* (because each participant made 14 judgments) and one for the specific *behavior* being judged (because each of the 28 behaviors took the form of a self, others, or threshold judgment).

We observed a significant Morality X Target interaction, $F(2, 2803.97) = 174.67, p < .001$, indicating that estimates varied by target. As in Study 1, we decomposed this interaction into a series of 2 X 2 interactions. We replicated our finding that participants thought they were more moral than others, $B = 19.81, SE = 1.11, t(2803.97) = 17.77, p < .001$. Of greater relevance was how self-judgments and others-judgments compared to the moral threshold. As in Study 1, we saw clear evidence of self-positivity: The self reported easily exceeding the moral threshold, $B = 14.46, SE = 1.07, t(2803.97) = 13.48, p < .001$. Once again, we saw weaker evidence of other-negativity: The others were believed to not quite live up to the moral standard reflected in the threshold, $B = -5.35, SE = 1.13, t(2803.97) = -4.76, p < .001$.

Unexpectedly, we also observed a Morality X Target X Frame interaction, $F(2, 2803.97) = 14.48, p < .001$. In light of this, we examined the robustness of self-positivity and other-negativity by analyzing moral behaviors and immoral behaviors separately for both the original and inverse frames. Speaking to the robustness of self-positivity, the self reported engaging in moral behaviors more frequently than the threshold and immoral behaviors less frequently than the threshold under both the original and inverse frame, *t*s > 2.99, *p*s < .003.

But like in Study 1, other-negativity was more finicky. In fact, it was only in one of four cases—when judging moral behaviors that took the inverse frame (i.e., how often one does *not* engage in an immoral behavior)—that evidence of other-negativity emerged, *B* = -14.26, *SE* = 2.55, *t*(507.72) = -5.59, *p* < .001. It is worth noting that other-negativity also emerged in Study 1 when judging these behaviors, but when they took their *original* frame (as immoral behaviors). But for that combination (immoral—original frame) as well as the other two (moral—original frame, immoral—inverse frame) in the present study, judgments of the others did not significantly differ from threshold, *t*s < 1.93, *p*s > .054 (see Table 2). In other words, the sporadic evidence for other-negativity—in contrast to the highly consistent evidence for self-positivity—seems not to identify a consistent, narrow context in which other-negativity emerges (given the inconsistencies in the Study 1 and 2 results), but instead reflects the relative flimsiness of these effects. Despite this inconsistent variation, other-negativity did emerge in the aggregate in both studies.

Study 3

After observing robust, consistent evidence of self-positivity and weaker evidence of other-negativity, Study 3 systematically varied the nature of the other being judged. This allowed us both to identify systematic variability in moral social perceptions and potentially to identify a feature of social targets that leads perceivers to switch from other-negativity to other-positivity.

Method

Participants and Design. Three hundred eighty-nine Americans (60.9% female, 38.8% male, 0.3% genderfluid; M_{age} = 38.95, SD_{age} = 12.81) recruited through Amazon's Mechanical Turk (AMT) took part in the study in exchange for nominal payment. Participants

Table 2

*Mean (*SD*) Estimates by Target, Frame, and Morality of Behaviors (Study 2)*

Frame **Morality**	**Self**	**Others**	**Threshold**
Original Frame			
Moral Behaviors	76.14_a (31.04)	56.05_b (26.23)	61.33_b (24.01)
Immoral Behaviors	18.91_a (25.67)	41.91_b (24.95)	44.35_b (25.34)
Morality Composite	57.23_a	14.14_b	16.98_b
Inverse Frame			
Moral Behaviors	71.07_a (32.50)	49.43_c (25.37)	63.70_b (25.41)
Immoral Behaviors	28.23_a (34.87)	42.76_b (24.34)	38.45_b (26.14)
Morality Composite	42.84_a	6.67_c	25.25_b
Overall			
Moral Behaviors	73.54_a (31.87)	52.49_c (25.95)	62.59_b (24.77)
Immoral Behaviors	23.69_a (31.06)	42.37_b (24.60)	41.21_b (25.91)
Morality Composite	49.85_a	10.12_c	21.38_b

Note. The morality composite reflects the moral behavior composite minus the immoral behavior composite. Means in the same row that do not share a subscript differ at the $p <$ .05 level. Moral and immoral behaviors of the inverse frame are the opposite of immoral and moral behaviors of the original frame, respectively.

made judgments about three of six possible targets. Two of these judgments were equivalent for everyone: *self* and *threshold.* But participants were randomly assigned to also make judgments about *others* that took one of four forms: *individuated, non-individuated, others, society*. Recall that participants in our first two studies judged only one target (self, others, or threshold). One strength of the between-participants designs of Studies 1 and 2 is they allowed us to consider how self and other judgments systematically depart from the moral threshold without offering any individual participant the opportunity to strategically sequence their three judgments in a preferred order. The complementary within-subjects design of Study 3 offers more power to examine how each of the four types of others (as well as the self) compare against participants' (own) ideographically assessed moral thresholds.

Procedure. To begin, all participants were told that they had been randomly assigned a participant code (53U7USS7P). At that point, they were asked to introduce themselves in three-to-four sentences: "You could say how you enjoy spending your time, what you do for a living, or anything else that would capture yourself in a few sentences." These two design features were of key relevance given the cover story later offered to those assigned to the individuated and non-individuated other conditions. Those participants would estimate the behaviors of another participant in the study identified only by their supposed participant code (non-individuated other) or their introductory remarks (individuated other).

At that point, participants learned they would be considering "a number of behaviors that might be performed in various situations." These behaviors composed a new set of (5) moral and (5) immoral actions (see Appendix B). Participants made three sets of judgments, each requiring them to consider all 10 behaviors. They completed the three sets in a random order; the order of the 10 behaviors was randomized within each set (though were segregated by morality when thresholds were estimated). For two of these three sets, participants indicated the percentage of the time that they engaged in each behavior (self judgments) and the behavioral rate that reflected the tipping point that differentiated morality from immorality for that particular behavior (threshold judgments). Participants' third set of judgments was about other people. The other target took one of four forms:

Individuated. For those assigned to the *individuated* condition, participants were told they would judge a randomly chosen other who had participated in the study previously.

Participants were allowed to read that person's introduction, composed by an actual person drawn from the same participant pool. The introductory remarks turned the target into an individuated other. In order to generate these introductions, we conducted a pretest (N = 89 Americans, AMT) in which we gave participants the same introduction prompt that those in the main study received. Three research assistants read every introduction and rated them on 5-point scales on three dimensions: the perceived *morality* of the person, how much *individuating* information was provided, and how *unusual* the content of the introduction was. We identified 7 descriptions that met three criteria: 1) the target's perceived morality was within 1 point of the sample mean, 2) the description was rated above-average in terms of individuation, and 3) the description was rated below-average in terms of unusualness. Each participant who was randomly assigned to rate an individuated target learned about the individuated other by reading a random 1 of these 7 introductions.

Non-individuated. Those assigned to the *non-individuated* condition were also told that they would judge a randomly chosen other who had participated in the study. But for these participants, they were informed only of this person's ID code, which took the same form as the self's code: "33A9p607g."

Others. Those assigned to the *others* condition were merely told they would judge the other participants in the study. In this way, they were estimating the percentage of the time not that a single person, but that the others considered in aggregate, engaged in each behavior. Notably, this condition matches the others condition used in the first two studies (but with a new set of behaviors).

Society. For those assigned to the *society* condition, they judged the percentage of time that others "in society in general" behaved in each way, in the situations described. This meant that these judgments were made about a collection of people, but not those who took part in the study.

Materials. For each of their three sets of judgments, participants considered the same 10 behaviors. These behaviors were identified by Galak and Critcher (2023) using two rounds of pretesting. This approach outsourced the generation of candidate behaviors to participants (to avoid experimenter bias) and then had a new group of participants indicate the extent to which each behavior resembled the prototype of an everyday moral or immoral behavior. From this two-stage process, the five moral and five immoral behaviors were identified.

Each behavior identified a specific action (e.g., help someone cross the street [e.g., elderly person, visually impaired person]) in a specified context (e.g., when [you; this person; participant 33A9p607g; they] observe[s] such a person in need). Participants estimated what percentage of the time the target engages in the behavior, or what percentage reflected the tipping-point threshold between morality and immorality. All such judgments were made on 0%-to-100% slider scales.

Results

We were again interested in testing for evidence of self-positivity. But given we varied the nature of the other target, we wanted to test whether (and when) other-negativity may turn into other-positivity. Toward this end, we again used a mixed model to predict participants' ratings. We included two fixed-effects predictors as well as their interaction. One was the *morality* of the particular behavior (+1 = moral, -1 = immoral). The second was a categorical variable for *target* (self, threshold, individuated other, non-individuated other, others, society). Finally, to account for the non-independence of participants' 10 judgments as well as different participants' ratings of the same behavior, we treated *participant* and *behavior* as random factors.

We observed a significant Morality X Target interaction, $F(5, 11262.20) = 104.28$, $p < .001$ (see Table 3), an omnibus test that would likely reflect that at least some targets deviated from the threshold. We then tested whether the self indicated that its own behaviors were more moral than were others'. Indeed, the self showed a better-than-others effect regardless of the nature of the other, *t*s > 5.13, *p*s < .001. To begin with immoral behaviors, the self indicated performing fewer immoral behaviors than any of the four other targets, *t*s > 4.54, *p*s < .001. Regarding moral behaviors, the self reported performing them more frequently than a non-individuated other, others in the study, or others in society, *t*s > 4.47, *p*s < .001, but no more often than an individuated other, *B* = 1.77, *SE* = 1.32, *t*(11537.62) = 1.35, *p* = .178.

We once again found evidence of self-positivity. That is, the 2(Morality) X 2(Target: self or threshold) interaction was significant, *t*(11262.20) = 15.39, *p* < .001. The self reported engaging in moral behaviors more often than their provided thresholds, *B* = 2.45, *SE* = 0.81, *t*(11262.20) = 3.04, *p* = .002. They also reported engaging in immoral behaviors less frequently than their reported thresholds, *B* = -15.11, *SE* = 0.81, *t*(11262.20) = -18.73, *p* < .001.

At this point, we turned to examining whether participants displayed evidence of other-negativity, or even other-positivity, depending on the form that that other took. In this case, the dividing line was clear. When people considered others as a collective—whether others in the study or society at large—participants displayed other-negativity: Participants believed that the behavior of both others in the study, *B* = -4.83, *SE* = 0.88, *t*(11262.20) = -5.48, *p* < .001, and others in society more generally, *B* = -6.99, *SE* = 0.88, *t*(11262.20) = -7.96, *p* < .001, fell short of the moral threshold. In contrast, when considering another as an individual—whether that person was individuated or not—the target was believed to exceed the moral threshold: Participants estimated that both an individuated other, *B* = 4.27, *SE* = 0.88, *t*(11262.19) = 4.86, *p* < .001, and a non-individuated other, *B* = 2.05, *SE* = 0.99,

Table 3

*Mean (*SD*) Estimates by Target and Morality of Behaviors (Study 3)*

Morality	**Self**	**Threshold**	**Others**			
			Individuated	**Non-individuated**	**Others**	**Society**
Moral Behaviors	63.85_{a} (31.83)	61.40_{b} (26.69)	61.98_{ab} (28.12)	58.32_{c} (28.23)	52.40_{d} (25.73)	50.55_{d} (26.77)
Immoral Behaviors	22.43_{a} (28.71)	37.54_{c} (30.87)	29.57_{b} (29.43)	30.36_{b} (28.39)	38.20_{cd} (26.87)	40.67_{d} (28.31)
Morality Composite	41.42_{a}	23.86_{c}	32.41_{b}	27.96_{b}	14.20_{d}	9.88_{d}

Note. The morality composite reflects the moral behavior composite minus the immoral behavior composite. Means in the same row that do not share the same subscript differ at the $p < .05$ level.

t(11262.19) = 2.07, *p* = .038, were better than threshold. The two targets that described collectives (others, society) did not significantly differ from each other, *B* = 2.16, *SE* = 1.11, *t*(11262.20) = 1.96, *p* = .051, nor did the two conditions that described individuals, *B* = 2.22, *SE* = 1.19, *t*(11262.20) = 1.86, *p* = .062. Because these differences were marginally significant, our final studies adopt the most conservative approach by comparing perceptions of a specific, non-individuated other in the study against perceptions of others in the study.

Study 4a

Study 3 showed that social targets' singularity—whether the target was individuated or not—elevated social perceptions of them above perceivers' moral thresholds. Studies 4a and 4b use a causal-chain design to probe a previously untested account for why assessments of individuals are elevated. Study 4a tests whether perceivers anticipate greater aversion from cynicism about individuals (vs. collectives). Study 4b manipulates these anticipated experiences to test their causal effect on social judgments.

Method

Participants and Design. Nine hundred eighty Americans who passed a captcha and a language-comprehension check were recruited through AMT. Fourteen participants did not pass a memory-based attention check at the study's conclusion. More specifically, they were unable to report that the study asked them to forecast how they would feel upon learning that they had overestimated or underestimated the prevalence of a behavior. Per our preregistered criterion, we excluded these participants from the main analyses, resulting in a final sample of 966 participants (70.7% female, 28.1% male, 1.2% non-binary; M_{age} = 39.52, SD_{age} = 12.72). Participants made judgments about one of two possible target others: *other* (a non-individuated individual) or *others* (the group of people from which the other would be drawn). The hypotheses, methods, sample size, exclusion criterion, and analysis plan were preregistered: https://aspredicted.org/1KF_BPF.

Procedure. All participants first received a participant ID code. This would be instrumental to the cover story in the *other* (*individual*) condition. Next, participants were told that they would consider "behaviors that people might engage in in different contexts." Those in the *others* condition were told they might estimate "how often the other participants in this study" engage in those behaviors. Those in the *other* condition instead learned they might estimate "how often one randomly chosen participant in this study" engages in the series of behaviors. Because Study 3 indicated that individuation was not necessary for the emergence of other-positivity, and thus to keep our design particularly conservative, the other was non-individuated and simply identified by their ostensible ID code: 33A9p607g.

All participants considered making a forecast about two behaviors. One was moral; the other, immoral. These behaviors were randomly selected, for each participant, from a set of 6 behaviors (3 moral, 3 immoral) that were used in Study 3 and, critically, would be used to complete the causal chain in Study 4b. Participants considered how they would feel if they overestimated or underestimated the actual percentage of the time the others (or the specific, non-individuated other) actually engaged in that behavior. When the error would reflect a cynical departure from the truth (underestimating a moral behavior or overestimating an immoral behavior), participants indicated how "guilty" and "mean" they would feel, on 7-point scales anchored at 1(*not at all*) and 7(*extremely*). These items were averaged into a *negative feelings* composite (r = .76). When the error would reflect an overly hopeful error (overestimating a moral behavior or underestimating an immoral behavior), participants indicated how "good" and "kind" they would feel. These too were expressed on 7-point scales anchored at 1(*not at all*) and 7(*extremely*). These items were averaged into a *positive feelings* composite (r = .88).

Results

We tested whether participants anticipated having stronger reactions to making errors about specific individuals as opposed to collectives. Toward this end, we conducted mixed models predicting anticipated negative feelings as well as anticipated positive feelings. These models included fixed-effects predictors of the *morality* of the behavior (+1 = moral, -1 = immoral) as well as the participant's *target* condition (+1 = other, -1 = others). The interaction term was included as well. To account for the non-independence of the judgments, we also included *participant* and the specific *behavior* being judged as random factors (see Table 4).

Anticipated negative feelings. In the first model, the focal effect of target was significant, $B = 0.21$, $SE = 0.04$, $t(962.45) = 4.83$, $p < .001$. This effect did not depend on whether the behavior was moral or immoral, $B = -0.03$, $SE = 0.03$, $t(960.62) = 1.07$, $p = .285$. The main effect reflected that perceivers anticipated feeling guiltier and meaner about being cynical about an individual other ($M = 3.52$) compared to a collective of others ($M = 3.10$). This suggests that differences in anticipated negative feelings about being overly cynical could explain why judgments of mere individuals are more positive than judgments of others

Table 4

*Mean (*SD*) Anticipated Feelings by Target and Morality of Behaviors (Study 4a)*

Anticipated feelings **Morality**	**Individual**	**Collective**
Positive Feelings		
Moral Behaviors	3.81 (1.65)$_a$	3.99 (1.74)$_a$
Immoral Behaviors	3.23 (1.60)$_a$	3.13 (1.65)$_a$
Negative Feelings		
Moral Behaviors	3.21 (1.55)$_a$	2.85 (1.61)$_b$
Immoral Behaviors	3.84 (1.78)$_a$	3.34 (1.77)$_b$

Note: Means in the same row that do not share the same subscript differ at the p < .05 level.

as a collective.

Anticipated positive feelings. In the second model, we found no effect of target, *B* = -0.02, *SE* = 0.04, *t*(962.78) = 0.48, *p* = .629. Though in this case there was a Target X Morality interaction, *B* = -0.07, *SE* = 0.03, *t*(960.04) = 2.09, *p* = .037. We proceeded to test for simple effects of target at each level of morality. Though notably, we did not find that perceivers expected to feel significantly more positively about making charitable errors about an individual other compared to others, whether that be in considering immoral behaviors, *B* = 0.05, *SE* = 0.05, *t*(1783.81) = 0.87, *p* = .386, or moral behaviors, *B* = -0.09, *SE* = 0.05, *t*(1784.62) = -1.64, *p* = .101. This does not support the possibility that differences in anticipated positive feelings about being unrealistically kind in one's assessments explain why judgments of individuals are more positive than judgments of collectives.

Study 4b

For Study 4b, we developed a manipulation that (as validated by a pretest) modified participants' beliefs about whether people tend to overestimate or underestimate just how guilty they will actually feel about cynical errors they make in social forecasting. After pretesting this intervention (to understand how it may operate differently with respect to individual or collective targets), we proceeded in the main study to test whether these anticipated negative feelings cause morally relevant judgments to become more positive.

Method

Participants and Design. One hundred ninety-nine undergraduate students from an American university took part in the main study in exchange for course credit. Per our preregistered criterion, we excluded 1 participant. This participant was unable to report at the study's conclusion what they were asked to contemplate as part of the study (correct answer: "why people may or may not feel guilt for overestimating how immoral others are"). This resulted in a final sample of 198 participants (58.1% female, 40.9% male, 1.0% who chose not to disclose; M_{age} = 20.85, SD_{age} = 2.01). Participants were randomly assigned to one of four conditions in a 2(target: other [individual] or others [collective]) X 2(guilt: overestimated or underestimated) full-factorial design. The hypotheses, methods, sample size, exclusion criterion, and analysis plan were preregistered: https://aspredicted.org/JYK_R5R.

Pretest. Prior to the main study, we developed a three-pronged manipulation that had the potential to change people's anticipated negative feelings about being cynical. We pretested the manipulation to understand whether it would indeed affect participants' anticipated negative feelings, but also to determine whether such influence would differ depending on the nature of the target (individual or collective). After all, if there is a natural aversion to being cynical about individuals that is not spontaneously present for collectives, then it may be easier to convince people that such negative feelings actually *would* arise when they are wrong about collectives than it would be to convince people that their anticipated aversion to cynicism toward an individual is incorrect. Understanding such nuances of how the manipulation operates would be crucial to form a preregistered prediction for the main study. The hypotheses, sample size, exclusion criterion, and analysis plan for the pretest were preregistered: https://aspredicted.org/blind.php?x=26M_NW6.

We began by explaining to participants (N = 383 Americans recruited from AMT, after excluding 11 participants who failed a preregistered attention check; 66.8% female, 32.1% male, 1.0% non-binary; M_{age} = 39.23, SD_{age} = 12.29) that "one common finding in social psychology is that people are pretty bad at estimating how they will feel if they misestimate how often others engage in morally relevant behaviors." For participants who underwent the *guilt underestimated* manipulation, they were told "people tend *not* to appreciate that they will feel guilt or that they were mean if they overestimate how often others engage in immoral behaviors (or underestimate how often others engage in moral behaviors.)" Those who considered the *guilt overestimated* manipulation were instead told that those who make these misforecasts "actually end up feeling less guilt or that they were not as mean as they anticipated."

To encourage internalization of this feedback, we asked participants to spend at least one minute writing on why people actually "feel pretty guilty or mean" (guilt underestimated condition) or "don't feel that guilty or mean" (guilt overestimated condition) when displaying cynicism. Finally, participants spent at least one more minute writing about why they think people "tend to mistakenly think they won't feel that guilty or mean" (guilt underestimated condition) or "tend to mistakenly think they will feel pretty guilty or mean" (guilt overestimated condition) when they engage in the relevant social misestimation. These manipulations were modified to apply to one of two *targets*: a randomly selected individual from the study with a specified ID code (other target condition) or others in the study (others target condition).

At that point, we showed participants two behaviors (1 moral, 1 immoral) that were randomly selected, for each participant, from the six that we used in Studies 4a-4b. For each behavior, participants indicated how *guilty* and *mean* they would feel if they underestimated (for the moral behavior) or overestimated (for the immoral behavior) the actual frequency with which the target engages in the behavior. This measure took the same form as in Study 4a. We averaged the two items to create an *anticipated negative feelings* composite for each behavior ($r = .78$).

We constructed a mixed model that included several fixed effects: guilt (+1 = underestimated, -1 = overestimated), target (+1 = other, -1 = others), and morality (+1 = moral, -1 = immoral). All interaction terms were included as well. To account for non-independence in the data, we also included random effects of participant and the specific behavior about which the forecast was made.

First, we observed a main effect of target, *B* = 0.16, *SE* = 0.08, *t*(378.32) = 2.04, *p* = .042, such that people anticipated feeling more guilt when being cynical about an individual as opposed to a collective. Though whether more guilt was anticipated depended on the nature of the guilt manipulation, *B* = 0.14, *SE* = 0.08, *t*(379.90) = 1.87, *p* = .062. Especially because this interaction fell just shy of significance, we were particularly interested in the specific patterns of simple effects that emerged, which would inform our preregistered hypotheses for the main study. When considering "others" as the target, participants led to embrace that it is common to underestimate how much guilt cynicism would inspire anticipated feeling worse about being cynical (*M* = 3.42) than those led to embrace that people tend to exaggerate how much guilt they will feel (*M* = 2.96), *B* = 0.23, *SE* = .11, *t*(378.02) = 2.20, *p* = .028. In contrast, the guilt manipulation failed to move participants' beliefs about how they would feel about being cynical toward an individual (*M*s = 3.48 and 3.51), *t* < 1.

These patterns informed a specific prediction that we preregistered for our main study: Encouraging participants to think that they would have a more aversive response to displaying cynicism would discourage such cynicism more for those judging others (but less so for a specific other).[1]

Procedure. The main study combined elements of Study 4a and the pretest. After being assigned a participant code (to help reinforce the cover story for the individual target condition), participants learned they would be making judgments about "some behaviors that people might engage in in different contexts." Before making those social judgments, participants completed one of the two anticipated guilt manipulations (overestimated or underestimated) validated in the pretest. Just like in the pretest, these manipulations took a slightly different form to match participants' target (other [individual] or others [collective])

[1] Note that for the present purposes, the mere presence of the asymmetry—regardless of why it occurs—is what is crucial. It may simply be that it is difficult to argue against the diagnosticity of an intuitive aversive response (as when one imagines victimizing an individual), whereas it is easier to rationally convince oneself that an emotionally muted intuition could be in error.

condition. At that point, participants made behavioral forecasts about the 6 behaviors (3 moral, 3 immoral) used in both Study 4a and the pretest. Depending on participants' target condition, they made these judgments about how often the "others in this study" (collective target condition) or "a randomly chosen participant, 33A9p607g" (individual target condition) engage in the behaviors in the situations described. At that point, participants also indicated—in a counterbalanced order—their moral thresholds and self-judgments as in Studies 1-3.[2]

Results

In order to complete the causal chain (and informed by the results of the pretest), we asked whether encouraging participants to think that people typically underestimate how aversive it would be to express cynicism would encourage more positive judgments about others (compared to about a specific other). To test this preregistered hypothesis, we used a model that was nearly identical to the one used in the pretest. Whereas the model was used to predict anticipated negative responses in the pretest, it predicted the social behavioral estimates here. Consistent with the central prediction, there was a significant Guilt X Target X Morality interaction, B = -2.35, SE = 0.71, t(982) = 3.31, p = .001.

We decompose this interaction in two complementary ways. To begin, we considered how the guilt manipulation affected others and an other separately, by decomposing the three-way interaction by target. The Guilt X Morality interaction was significant for both the individual other and others, but the direction of each interaction was different. As predicted, an encouragement to embrace that participants would likely underestimate their feelings of guilt (thereby increasing the anticipated feelings) led to more positive estimates of others, B = 2.28, SE = 0.99, t(982) = 2.30, p = .022. Not only did the manipulation not have an analogous

[2] Given we expected the guilt manipulations would have an effect on the social judgments, our preregistered prediction was only that we would replicate our finding that self-judgments would exceed the moral threshold. The Supplemental Materials includes this preregistered test, which was confirmed, as well as additional non-preregistered analyses showing that other-negativity emerged only when the target was a collective (others) and participants were led to embrace that cynicism-related guilt is typically overestimated.

effect on judgments of a specific other (as foreshadowed by the pretest), but it actually (unexpectedly) had a reverse effect, *B* = -2.42, *SE* = 1.02, *t*(982) = -2.37, *p* = .018.

Next, we decompose the three-way interaction in a complementary way. More specifically, we considered how moral others (a collective) as opposed to an other (an individual) were judged following each guilt manipulation. Keep in mind that the pretest showed that it was only when participants were led to think that they would *underestimate* how much guilt cynicism would inspire that we essentially eliminated the other-others asymmetry in anticipated negative feelings. When participants were encouraged to embrace that people tend to overestimate how much guilt cynicism would invite, we found our oft-observed finding that others were judged more negatively than an individual other, *B* = 5.14, *SE* = 1.01, *t*(982) = 5.11, *p* < .001. But when anticipated guilt was manipulated to be high (by convincing people that such estimates tend to be *underestimated*), others and a specific other were not judged differently, *B* = 0.45, *SE* = 1.00, *t*(982) = 0.44, *p* = .657 (see Table 5).

These results thus complete the causal chain. By manipulating how much guilt people anticipate should they be cynical (a manipulation our pretest showed affects those considering judgments of others, but not a specific other), we were able to predictably eliminate the effect that people are more positive in their moral forecasts of a specific, individual other instead of a collective of others.

General Discussion

Social psychologists have long appreciated the special status the self holds in its own mind. When most people see themselves as better than most people, we can be confident that the median person tends to engage in overplacement, a form of overconfidence (Moore & Healy, 2008). The better-than-average effect is the most famous form of this bias (Zell et al., 2020). In such demonstrations, social perceptions mostly serve as the comparison standard by which to identify typical self-perceptions as inflated.

Table 5

*Mean (*SD*) Estimates by Guilt, Target, and Morality of Behaviors (Study 4b)*

Morality		
Guilt	**Other (Individual)**	**Others (Collective)**
Moral Behaviors		
Overestimated	51.33_{a} (28.14)	44.91_{a} (27.01)
Underestimated	46.06_{a} (27.91)	47.09_{a} (25.62)
Immoral Behaviors		
Overestimated	22.11_{a} (22.19)	36.27_{b} (26.14)
Underestimated	26.54_{a} (26.14)	29.34_{a} (26.07)
Morality Composite		
Overestimated	29.22_{a}	8.64_{b}
Underestimated	19.52_{a}	17.75_{a}

Note. The morality composite reflects the moral behavior composite minus the immoral behavior composite. Means in the same row that do not share a subscript differ at the $p < .05$ level.

Despite a decades-long focus on this phenomenon, the present paper addressed two major questions that this literature had yet to tackle. First, we introduced a new measure—the moral threshold—that allowed us to determine whether the self and others are viewed as morally adequate or inadequate. We found consistent evidence—across different countries, using both online and in-lab samples, and across several sets of behaviors—of self-positivity. Other people were instead seen as falling short of this moral threshold, though these effects were smaller and less robust.

Second, we noted that self-other comparisons often treat the “other” as something of an afterthought, a comparative standard that makes self-judgments more interpretable. We instead systematically explored how different social targets stack up against the moral threshold as well as each other. Whereas collectives were seen to fall short of the moral threshold, specific individuals—even about whom no individuating information was provided—were seen to exceed it. A final pair of studies helped to explain the counternormative finding that randomly selected individuals are judged more positively than the group of individuals from which they are drawn. Perceivers anticipated feeling worse about cynically misjudging a person than they did about cynically misjudging people. By experimentally manipulating those anticipations, the judgment gap evaporated.

The Assessment of Limitations (Table 6) considers issues for which future research will be necessary. Methodologically, these include the need to extend future investigations beyond two cultural contexts, expand the number of morally relevant behaviors considered beyond 24, and consider whether inconsistencies in which behaviors led to others-negativity are explained by random variation or population characteristics. In terms of research focus, future research will be necessary to better understand: how perceivers’ moral thresholds operate in drawing moral character inferences from behavioral information, the origin of moral thresholds, and whether stereotyping-and-prejudice findings attributed to individuation are explained by individuating information or instead mere individuation (as the present results may suggest).

Theoretical Implications and Open Questions for Future Research

Moral threshold. In other literatures, behavioral scientists have long recognized the benefit of identifying neutral reference points that serve as a baseline for understanding other

Table 6

Assessment of Limitations

Dimension	Assessment
Diversity of samples	Although studies made use of both college undergraduate and online samples, these were drawn from only two countries: one in North America and one in Europe.
Diversity of materials	Across the studies, we drew on 12 everyday moral and 12 everyday immoral behaviors. Although some of these were selected using a procedure to avoid experimenter bias, this does not guarantee that the results generalize to all moral and immoral behaviors.
Robustness of self-positivity and others-negativity	Across studies, self-positivity was robust. Although we always observed evidence of others-negativity in the aggregate, which behaviors drove others-negativity sometimes varied across studies. Whether this is attributable to random variation or specific characteristics of the populations sampled is currently unclear.
Open questions regarding how perceivers integrate knowledge of, or estimates about, behaviors to arrive at global moral character evaluations	The present paper lays a foundation for what we hope will be future developments in the study of self and social perception. Natural next questions include whether deviations from moral neutrality (the threshold) influence broader moral character assessments differently depending on whether such deviations are positive or negative (see also Klein & Epley, 2017), whether these deviations from moral neutrality are mapped onto moral character evaluations differently for self and social perception, and the process by which people

	synthesize evaluations based on different morally relevant behaviors (some of which may exceed while others fall short of the threshold) into a more global perception of a person. We hope these possibilities can breathe new life into a question that is essentially settled—that most people see themselves as better than most others on most dimensions of personal import (Zell et al., 2020). The present efforts that identified self-perceptions as exceeding people's moral threshold and social perceptions as variable (depending on the status of the moral target as an individual) offer a qualitatively new foundation that could fuel a resurgence of interest in these topics.
Origin of moral thresholds	One open question is what influences the formation of the moral threshold. Whereas previous research has shown that people use their own personal standing as a reference point against which they judge others (Dunning & Hayes, 1996), our introduction of the moral threshold enables the identification of a reference point that allows both self and social judgments to become more interpretable. That said, much as Dunning and Hayes's (1996) work might suggest, one possibility is that self-views inform the creation of (a typically less stringent) perceived threshold. Whether perceptions of social targets—either individuals or collectives—may also inform moral thresholds is itself a possibility that awaits direct test. As reported in the Supplemental Materials (Tables S1-S6), moral thresholds were correlated with both self and social judgments.

Implications for the importance of individuation in stereotyping and prejudice effects remain to be explored	The stereotyping-and-prejudice literature has long argued for the importance of individuation as a tool to reduce stereotyping and discrimination (Rubinstein et al., 2018; Sherman et al., 2005). The present results suggest that more positive impressions of individuated targets may be less attributable to individuating information and more a function of their merely being individuals. If so, the present results imply that altering perceivers' orientations toward disfavored groups may benefit from encouraging construals of such targets as specific (even non-individuated) people, not collectives. This insight may prove most useful when direct contact, which is known to carry intergroup benefits (Allport, 1954; Pettigrew & Tropp, 2006; Paluck et al., 2019), is not feasible.

judgments and outcomes. As one example, one of prospect theory's major innovations was to recognize that monetary outcomes acquire psychological significance based on how they compare to a neutral reference point (Kahneman & Tversky, 1979). Identification of this reference point allowed for the discovery that outcomes possess different meaning depending on their distance, as well as the directionality of the deviation, from the neutral baseline. Relatedly, attitudes researchers have appreciated the importance of identifying people's basic evaluations not merely on an ordinal scale, but by identifying whether such reactions are fundamentally positive or negative (Cacioppo et al., 1997, 1999). Identifying attitudes as mostly positive or negative allows for the recognition of what targets will be approached as opposed to avoided, a fundamental functional purpose of the preceding evaluations (e.g., Fazio et al., 2004; Shook et al., 2007).

Our introduction of the moral threshold measure is not the first effort to identify standards by which people are considered. For example, research rooted in self-discrepancy theory has developed measures to assess people's ideal and ought selves (on the positive end; Higgins, 1987) and one's feared self (on the negative end; Carver et al., 1999; Markus & Nurius, 1986). Such measures proved their practical utility because deviations from such standards have predictable consequences—for example, guilt or disappointment (e.g., Carver et al., 1999; Tangney et al., 1998). Our threshold measure distinguishes itself by aiming to capture not an extreme (positive or negative) version of a specific target, but instead a neutral threshold that can be applied across targets. People may have different aspirations for themselves than they do for others, but the moral threshold measure identifies a common benchmark by which those different perceptions and aspirations can be understood. Future research may explore the combined influence of exceeding certain standards (e.g., the moral threshold) while falling short of others (e.g., the ideal self). One possibility is people who fall short of their idiosyncratic standards (e.g., their ideal self) may take comfort upon considering they have exceeded their moral thresholds, a strategy that those who are particularly well-adjusted may even employ spontaneously.

Finally, future research may benefit from extending our threshold concept to non-moral domains. After all, much better-than-average work reveals that the self sees itself as more competent than others. Identifying the thresholds that differentiate competence from incompetence would allow for tests of whether the self and others are seen to be fundamentally competent or incompetent, possibly again to show that specific individuals are viewed as competent even as the groups from which they are drawn are dismissed as incompetent. Although some mechanisms that have been proposed to explain why specific others are viewed more positively than the collective of others from which they are drawn apply specifically to the moral domain (Critcher & Dunning, 2013, 2014), the mechanism identified in the present work may extend to non-moral domains as well.

People are bad, a person is good. The present work also provides the first empirical test of Critcher and Dunning's (2014) theoretical proposal that perceivers may be reluctant to be harsh in their assessments of specific individuals because such cynical errors would be interpreted as guilt-inspiring acts of meanness or aggression. Although we directly documented this psychological process, a deeper mechanistic question can be proposed about its origin. Critcher and Dunning (2014) argued that personal relationships are forged with specific individuals, not the broader collectives from which they hail. As such, people may be practiced at not preemptively dismissing the moral worthiness of specific people, lest they reject potential social relationships before such targets have even had a chance to prove trustworthy. After all, smooth social functioning is facilitated by adherence to a basic norm of trust and respect (Dunning et al., 2014), especially toward individuals.

From this perspective, our results can be interpreted through the lens of error management theory (Haselton & Buss, 2000), which recognizes that certain errors are costlier than others; thus, people prefer to err in a certain (less costly) direction. Typically, this theory is appealed to to potentially explain why errors occur in a certain direction (e.g., Johnson et al., 2013; Haselton & Nettle, 2006). We go beyond making the functional argument to identify the more proximal psychological mechanism that drives this potentially functional bias: different anticipated negative experiences following a directional error. At least speculatively, we can delve one step deeper to consider why (beyond merely noting its potential functionality) the anticipated negative experience of being cynical about individual others is especially strong. When one is wrong about a specific individual, there is no ambiguity regarding who is being misjudged. There is a specific victim. In contrast, when one is wrong about the prevalence of a behavior in a collective or population, there is ambiguity about *which* targets are being misjudged; after all, the judgment is not meant to apply to everyone. One intriguing implication of this logic is if it were known that all members of a population had behaved the same way (meaning that an error in the population judgment would imply a misjudgment of each individual member of a population), more positive judgments of the population might result. By analogy, consider someone trying to guess the average age of everyone in a room. We suspect they would drop the typical impulse—when guessing a specific adult's age—to err young when describing the collective. But if one walked into a high school reunion of unknown graduation year—in which all alumni were the same age—we suspect that impulse might return.

Another open question is whether people are correct in anticipating how guilty they will feel about their cynicism toward individuals (versus collectives). On the one hand, previous research suggests that individual victims do elicit stronger negative responses than collections of victims (Kogut & Ritov, 2005a, 2005b). On the other hand, the affective forecasting literature shows that people tend to overestimate both the intensity and the longevity of their affective responses (Dunn et al., 2003; Finkenauer et al., 2007; Gilbert et al., 1998; Wilson & Gilbert, 2005; Wilson et al., 2000). Note that in the present work, participants would not have an obvious opportunity to have their cynicism confirmed, but anticipated aversive feelings still causally affected forecasts. This may speak to how well-engrained the process we identified is, automatically generalizing to contexts in which forecasters will not even have the opportunity to have their estimates confirmed or discredited.

Conclusion

There is an interpretation gap between behavioral estimates of the self and others and the evaluations that such beliefs imply. Merely measuring self and social perceptions allows for identifying systematic differences between the two (e.g., the better-than-average effect). At least in the moral domain, the better-than-average effect robustly reflects self-positivity, but emerges alongside both other-positivity and other-negativity, depending on the nature of the other. The self can thus avoid the special moral guilt of expressing cynicism about specific others while maintaining a sense of their exceptional standing in the population at large. Such magnanimity may even contribute to the self's sense of its own moral superiority.

References

Alicke, M. D., Klotz, M. L., Breitenbecher, D. L., Yurak, T. J., & Vredenburg, D. S. (1995). Personal contact, individuation, and the better-than-average effect. *Journal of Personality and Social Psychology, 68*(5), 804–825. https://doi.org/10.1037/0022-3514.68.5.804

Allport, G. (1954), *The nature of prejudice*. Basic Books.

Balcetis, E. & Dunning, D. (2013). Considering the situation: Why people are better social psychologists than self-psychologists. *Self and Identity*, *12*(1), 1-15. https://doi.org/10.1080/15298868.2011.617886

Brown, J. D. (1986). Evaluations of self and others: Self-enhancement biases in social judgments. *Social Cognition*, *4*(4), 353-376. https://doi.org/10.1521/soco.1986.4.4.353

Brown, J. D. (2012). Understanding the better than average effect: Motives (still) matter. *Personality and Social Psychology Bulletin, 38*(2), 209-219. https://doi.org/10.1177/0146167211432763

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1997). Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space. *Personality and Social Psychology Review*, *1*(1), 3-25. https://doi.org/10.1207/s15327957pspr0101_2

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1999). The affect system has parallel and integrative processing components: Form follows function. *Journal of Personality and Social Psychology, 76*(5), 839-855. https://doi.org/10.1037/0022-3514.76.5.839

Carlson, R. W., Maréchal, M. A., Oud, B., Fehr, E., & Crockett, M. J. (2020). Motivated misremembering of selfish decisions. *Nature Communications*, *11*(1), 2100. https://doi.org/10.1038/s41467-020-15602-4

Carver, C. S., Lawrence, J. W., & Scheier, M. F. (1999). Self-discrepancies and affect: Incorporating the role of feared selves. *Personality and Social Psychology Bulletin*, *25*(7), 783-792. https://doi.org/10.1177/0146167299025007002

Critcher, C. R. & Dunning, D. (2011). No good deed goes unquestioned: Cynical reconstruals maintain belief in the power of self-interest. *Journal of Experimental Social Psychology*, *47*(6), 1207-1213. https://doi.org/10.1016/j.jesp.2011.05.001

Critcher, C. R. & Dunning, D. (2013). Predicting persons' versus a person's goodness: Behavioral forecasts diverge for individuals versus populations. *Journal of Personality and Social Psychology, 104*(1), 28–44. https://doi.org/10.1037/a0030836

Critcher, C. R. & Dunning, D. (2014). Thinking about others versus another: Three reasons judgments about collectives and individuals differ. *Social and Personality Psychology Compass*, *8*(12), 687-698. https://doi.org/10.1111/spc3.12142

Critcher, C. R., Helzer, E. G., & Dunning, D. (2011). Self-enhancement via redefinition: Defining social concepts to ensure positive views of self. In M. D. Alicke & C. Sedikides (Eds.), *Handbook of self-enhancement and self-protection* (pp. 69-91). The Guilford Press.

Critcher, C. R., Inbar, Y., & Pizarro, D. A. (2013). How quick decisions illuminate moral character. *Social Psychological and Personality Science*, *4*(3), 308-315. https://doi.org/10.1177/1948550612457688

Dunn, E. W., Wilson, T. D., & Gilbert, D. T. (2003). Location, location, location: The misprediction of satisfaction in housing lotteries. *Personality and Social Psychology Bulletin*, *29*(11), 1421-1432. https://doi.org/10.1177/0146167203256867

Dunning, D. & Hayes, A. F. (1996). Evidence for egocentric comparison in social judgment. *Journal of Personality and Social Psychology, 71*(2), 213–229. https://doi.org/10.1037/0022-3514.71.2.213

Dunning, D., Anderson, J. E., Schlösser, T., Ehlebracht, D., & Fetchenhauer, D. (2014). Trust at zero acquaintance: More a matter of respect than expectation of reward. *Journal of Personality and Social Psychology, 107*(1), 122–141. https://doi.org/10.1037/a0036673

Epley, N. & Dunning, D. (2000). Feeling "holier than thou": Are self-serving assessments produced by errors in self- or social prediction? *Journal of Personality and Social Psychology, 79*(6), 861–875. https://doi.org/10.1037/0022-3514.79.6.861

Epley, N. & Dunning, D. (2006). The mixed blessings of self-knowledge in behavioral prediction: Enhanced discrimination but exacerbated bias. *Personality and Social Psychology Bulletin*, *32*(5), 641-655. https://doi.org/10.1177/0146167205284007

Everett, J. A. C., Pizarro, D. A., & Crockett, M. J. (2016). Inference of trustworthiness from intuitive moral judgments. *Journal of Experimental Psychology: General, 145*(6), 772–787. https://doi.org/10.1037/xge0000165

Fazio, R. H., Eiser, J. R., & Shook, N. J. (2004). Attitude formation through exploration: Valence asymmetries. *Journal of Personality and Social Psychology, 87*(3), 293–311. https://doi.org/10.1037/0022-3514.87.3.293

Fetchenhauer, D. & Dunning, D. (2010). Why so cynical? Asymmetric feedback underlies misguided skepticism regarding the trustworthiness of others. *Psychological Science*, *21*(2), 189-193. https://doi.org/10.1177/0956797609358586

Finkenauer, C., Gallucci, M., van Dijk, W. W., & Pollmann, M. (2007). Investigating the role of time in affective forecasting: Temporal influences on forecasting accuracy. *Personality and Social Psychology Bulletin, 33*(8), 1152-1166. https://doi.org/10.1177/0146167207303021

Galak, J. & Critcher, C. R. (2023). Who sees which political falsehoods as more acceptable and why: A new look at in-group loyalty and trustworthiness. *Journal of Personality and Social Psychology, 124*(3), 593–619. https://doi.org/10.1037/pspi0000264

Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology, 75*(3), 617–638. https://doi.org/10.1037/0022-3514.75.3.617

Golebiowska, E. A. (2001). Individual-targeted tolerance and timing of group membership disclosure. *The Journal of Politics*, *63*(4), 1017-1040. https://doi.org/10.1111/0022-3816.00099

Haselton, M. G. & Buss, D. M. (2000). Error management theory: A new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology, 78*(1), 81–91. https://doi.org/10.1037/0022-3514.78.1.81

Haselton, M. G. & Nettle, D. (2006). The paranoid optimist: An integrative evolutionary model of cognitive biases. *Personality and Social Psychology Review, 10*(1), 47-66. https://doi.org/10.1207/s15327957pspr1001_3

Helzer, E. G. & Dunning, D. (2012). Why and when peer prediction is superior to self-prediction: The weight given to future aspiration versus past achievement. *Journal of Personality and Social Psychology, 103*(1), 38–53. https://doi.org/10.1037/a0028124

Hester, N., Payne, B. K., & Gray, K. (2020). Promiscuous condemnation: People assume ambiguous actions are immoral. *Journal of Experimental Social Psychology*, *86*, 103910. https://doi.org/10.1016/j.jesp.2019.103910

Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review, 94*(3), 319–340. https://doi.org/10.1037/0033-295X.94.3.319

Horswill, M. S., Waylen, A. E., & Tofield M. I. (2004). Drivers' ratings of different components of their own driving skill: A greater illusion of superiority for skills that relate to accident involvement. *Journal of Applied Social Psychology, 34*(1), 177-195. https://doi.org/10.1111/j.1559-1816.2004.tb02543.x

Johnson, D. D., Blumstein, D. T., Fowler, J. H., & Haselton, M. G. (2013). The evolution of error: Error management, cognitive constraints, and adaptive decision-making biases. *Trends in Ecology & Evolution*, *28*(8), 474-481. https://doi.org/10.1016/j.tree.2013.05.014

Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263-292. https://doi.org/10.2307/1914185

Klar, Y. & Giladi, E. E. (1999). Are most people happier than their peers, or are they just happy? *Personality and Social Psychology Bulletin, 25*(5), 586-595. https://doi.org/10.1177/0146167299025005004

Klein, N. & Epley, N. (2016). Maybe holier, but definitely less evil, than you: Bounded self-righteousness in social judgment. *Journal of Personality and Social Psychology, 110*(5), 660–674. https://doi.org/10.1037/pspa0000050

Klein, N. & Epley, N. (2017). Less evil than you: Bounded self-righteousness in character inferences, emotional reactions, and behavioral extremes. *Personality and Social Psychology Bulletin*, *43*(8), 1202-1212. https://doi.org/10.1177/0146167217711918

Kogut, T. & Ritov, I. (2005a). The “identified victim” effect: An identified group, or just a single individual?. *Journal of Behavioral Decision Making*, *18*(3), 157-167. https://doi.org/10.1002/bdm.492

Kogut, T. & Ritov, I. (2005b). The singularity effect of identified victims in separate and joint evaluations. *Organizational Behavior and Human Decision Processes*, *97*(2), 106-116. https://doi.org/10.1016/j.obhdp.2005.02.003

Kruger, J. & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Kruger, J., & Gilovich, T. (1999). "Naive cynicism" in everyday theories of responsibility assessment: On biased assumptions of bias. *Journal of Personality and Social Psychology, 76*(5), 743–753. https://doi.org/10.1037/0022-3514.76.5.743

Kruger, J. & Gilovich, T. (2004). Actions, intentions, and self-assessment: The road to self-enhancement is paved with good intentions. *Personality and Social Psychology Bulletin*, *30*(3), 328-339. https://doi.org/10.1177/0146167203259932

Kruger, J. & Savitsky, K., (2009). On the genesis of inflated (and deflated) judgments of responsibility. *Organizational Behavior and Human Decision Processes, 108*(1), 143-152. https://doi.org/10.1016/j.obhdp.2008.06.002

Kruger, J. (1999). Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology, 77*(2), 221–232. https://doi.org/10.1037/0022-3514.77.2.221Justin

Kruger, J., Windschitl, P. D., Burrus, J., Fessel, F., & Chambers, J. R. (2008). The rational side of egocentrism in social comparisons. *Journal of Experimental Social Psychology*, *44*(2), 220-232. https://doi.org/10.1016/j.jesp.2007.04.001

Lammers, J., Pauels, E., Fleischmann, A., & Galinsky, A. D. (2022). Why people hate congress but love their own congressperson: An information processing explanation. *Personality and Social Psychology Bulletin*, *48*(3), 412-425. https://doi.org/10.1177/01461672211002336

Markus, H. & Nurius, P. (1986). Possible selves. *American Psychologist, 41*(9), 954–969. https://doi.org/10.1037/0003-066X.41.9.954

Messick, D. M., Bloom, S., Boldizar, J. P., & Samuelson, C. D. (1985). Why we are fairer than others. *Journal of Experimental Social Psychology*, *21*(5), 480-500. https://doi.org/10.1016/0022-1031(85)90031-9

Miller, D. T. & Ratner, R. K. (1998). The disparity between the actual and assumed power of self-interest. *Journal of Personality and Social Psychology, 74*(1), 53–62. https://doi.org/10.1037/0022-3514.74.1.53

Miller, D. T. (1999). The norm of self-interest. *American Psychologist, 54*(12), 1053–1060. https://doi.org/10.1037/0003-066X.54.12.1053

Moore, D. A. & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115*(2), 502–517. https://doi.org/10.1037/0033-295X.115.2.502

Morewedge, C. K. (2009). Negativity bias in attribution of external agency. *Journal of Experimental Psychology: General, 138*(4), 535–545. https://doi.org/10.1037/a0016796

Murray, S. L. (1999). The quest for conviction: Motivated cognition in romantic relationships. *Psychological Inquiry, 10*(1), 23-34.

Paluck, E. L., Green, S. A., & Green, D. P. (2019). The contact hypothesis re-evaluated. *Behavioural Public Policy, 3(2)*, 129-158.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology, 90(5),* 751–783.

Pizarro, D. A. & Tannenbaum, D. (2012). Bringing character back: How the motivation to evaluate character influences judgments of moral blame. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 91–108). American Psychological Association. https://doi.org/10.1037/13091-005

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, *28*(3), 369-381. https://doi.org/10.1177/0146167202286008

Rubinstein, R. S., Jussim, L., & Stevens, S. T. (2018). Reliance on individuating information and stereotypes in implicit and explicit person perception. *Journal of Experimental Social Psychology*, *75*, 54-70. https://doi.org/10.1016/j.jesp.2017.11.009

Sears, D. O. (1983). The person-positivity bias. *Journal of Personality and Social Psychology, 44*(2), 233–250. https://doi.org/10.1037/0022-3514.44.2.233

Shah, J., Higgins, T., & Friedman, R. S. (1998). Performance incentives and means: How regulatory focus influences goal attainment. *Journal of Personality and Social Psychology, 74*(2), 285–293. https://doi.org/10.1037/0022-3514.74.2.285

Sherman, J. W., Stroessner, S. J., Conrey, F. R., & Azam, O. A. (2005). Prejudice and stereotype maintenance processes: Attention, attribution, and individuation. *Journal of Personality and Social Psychology, 89*(4), 607–622. https://doi.org/10.1037/0022-3514.89.4.607

Shook, N. J., Fazio, R. H., & Eiser, J. R. (2007). Attitude generalization: Similarity, valence, and extremity. *Journal of Experimental Social Psychology*, *43*(4), 641-647. https://doi.org/10.1016/j.jesp.2006.06.005

Small, D. A. & Loewenstein, G. (2003). Helping a victim or helping the victim: Altruism and identifiability. *Journal of Risk and uncertainty*, *26*, 5-16. https://doi.org/10.1023/A:1022299422219

Small, D. A., Loewenstein, G., & Slovic, P. (2007) Sympathy and callousness: The impact of deliberative thought on donations to identifiable and statistical victims. *Organizational Behavior and Human Decision Processes, 102*(2), 143-153. https://doi.org/10.1016/j.obhdp.2006.01.005

Steimer, A. & Mata, A. (2016). Motivated implicit theories of personality: My weaknesses will go away, but my strengths are here to stay. *Personality and Social Psychology Bulletin*, *42*(4), 415-429. https://doi.org/10.1177/0146167216629437

Tangney, J. P., Niedenthal, P. M., Covert, M. V., & Barlow, D. H. (1998). Are shame and guilt related to distinct self-discrepancies? A test of Higgins's (1987) hypotheses. *Journal of Personality and Social Psychology, 75*(1), 256–268. https://doi.org/10.1037/0022-3514.75.1.256

Tappin, B. M. & McKay, R. T. (2016). The illusion of moral superiority. *Social Psychological and Personality Science*, *8*(6), 623-631. https://doi.org/10.1177/1948550616673878

Van Lange, P. A. & Sedikides, C. (1998). Being more honest but not necessarily more intelligent than others: Generality and explanations for the Muhammad Ali effect. *European Journal of Social Psychology*, *28*(4), 675-680. https://doi.org/10.1002/(SICI)1099-0992(199807/08)28:4%3C675::AID-EJSP883%3E3.0.CO;2-5

Walker, J. & Gilovich, T. (2021). The streaking star effect: Why people want superior performance by individuals to continue more than identical performance by groups. *Journal of Personality and Social Psychology, 120*(3), 559–575. https://doi.org/10.1037/pspa0000256

Wenzel, M. (2005). Misperceptions of social norms about tax compliance: From theory to intervention. *Journal of Economic Psychology*, *26*(6), 862-883. https://doi.org/10.1016/j.joep.2005.02.002

Williams, E. F. & Gilovich, T. (2012). The better-than-my-average effect: The relative impact of peak and average performances in assessments of the self and others. *Journal of*

Experimental Social Psychology, *48*(2), 556-561. https://doi.org/10.1016/j.jesp.2011.11.010

Wilson, T. D. & Gilbert, D. T. (2005). Affective forecasting: Knowing what to want. *Current Directions in Psychological Science*, *14*(3), 131-134. https://doi.org/10.1111/j.0963-7214.2005.00355.x

Wilson, T. D., Wheatley, T., Meyers, J. M., Gilbert, D. T., & Axsom, D. (2000). Focalism: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology, 78*(5), 821–836. https://doi.org/10.1037/0022-3514.78.5.821

Zell, E. & Alicke, M. D. (2011). Age and the better-than-average effect. *Journal of Applied Social Psychology*, *41*(5), 1175-1188. https://doi.org/10.1111/j.1559-1816.2011.00752.x

Zell, E., Strickhouser, J. E., Sedikides, C., & Alicke, M. D. (2020). The better-than-average effect in comparative self-evaluation: A comprehensive review and meta-analysis. *Psychological Bulletin, 146*(2), 118–149. https://doi.org/10.1037/bul0000218

Appendix A – Moral and Immoral Behaviors, Studies 1-2

Moral Behaviors

- Admitting one is wrong and apologizing, when having an argument and realizing one is wrong
- Correcting the cashier and returning the extra money, when one receives too much change after paying at a store or café
- Flushing the toilet, after one uses a public restroom
- Giving up a bus or subway seat to someone in need (e.g., an old, disabled, or pregnant person), when there are more passengers than available seats
- Recycling trash that is recyclable
- Voting, when there is an election
- Washing one's hands, after one uses the restroom

Immoral Behaviors

- Acting passively (not intervening or saying something), when one witnesses an injustice (e.g., someone making a racist comment at someone else, someone physically or verbally assaulting their partner)
- Communicating in a rude (and not a kind) way, when interacting with a service employee (e.g., a table server, a cashier, a public employee) and one is frustrated with something
- Getting aggressive and raising one's voice or shouting at another person, when one is discussing politics or some other controversial subject with someone
- Gossiping about or making fun of others behind their backs, when one is engaged in day-to-day conversation
- Looking at someone's answers or cheating, when one is taking a test or exam

- Saying one is too busy or giving some other excuse for not helping, when someone asks for help with something that would take less than 5 minutes to do
- Throwing trash on the ground, when one has trash and there are no trashcans nearby

Appendix B – Moral and Immoral behaviors, Studies 3-4

The behaviors kept in Studies 4a and 4b are marked with an asterisk.

Moral Behaviors

- *Returning a lost item (e.g., by tracking down the owner, to a "lost and found") when finding one (worth $20 or more)
- *Volunteering to give up one’s seat so others can sit together (when seeing two people struggling to find seats together)
- *Returning excess change to a cashier (when in this situation as a costumer)
- Helping a stranger who dropped possessions s/he was carrying to retrieve them (when seeing this happen)
- Helping someone cross the street (e.g., elderly person, visually impaired person), when one observes such a person in need

Immoral Behaviors

- Sharing with someone else a secret that one was asked to keep (thereby going against the person's wishes)
- *Making a racist joke
- *Knowingly lying on one’s tax returns
- Making fun of someone in front of other people
- *Pretending not to hear when one hears someone calling for help

Supplemental Materials

Correlations between Self, Social, and Threshold Judgments: Studies 3 and 4b

We were interested in how the target judgments—self, other, and threshold—were correlated. Studies 3 and 4b permit these tests. Of course, some behaviors may simply invite higher judgments (across all three targets) than do others. This could artificially increase correlations. To avoid the influence of this extraneous source of variation, we first standardized the estimates across all targets, but for each behavior separately. In other words, we calculated the grand mean and standard deviation of the self, other, and threshold judgments for each behavior and used those to Z-score each individual judgment. This meant that the mean and standard deviation of all judgments for a specific behavior were 0 and 1, respectively, even though the mean and standard deviations of self, other, and threshold judgments individually would almost certainly depart from these values. The correlations from Study 3 are in Tables S1-S3; those for Study 4b are in Tables S4-S6.

Table S1

Correlations Between Target Judgments (Study 3)

Target	Threshold	Individuated	Non-individuated	Others	Society
Self	.41***	.46***	.51***	.45***	.43***
Threshold	-	.35***	.49***	.43***	.45***

Note. ***p < .001.

Table S2

Correlations Between Target Judgments, for Moral Behaviors (Study 3)

Target	Threshold	Individuated	Non-individuated	Others	Society
Self	.40***	.36***	.40***	.51***	.57***
Threshold	-	.28***	.40***	.38***	.47***

Note. ***p < .001.

Table S3

Correlations Between Target Judgments, for Immoral Behaviors (Study 3)

Target	Threshold	Individuated	Non-individuated	Others	Society
Self	.48***	.57***	.67***	.57***	.49***
Threshold	-	.42***	.60***	.46***	.45***

Note. ***p < .001.

Table S4

Correlations Between Target Judgments (Study 4b)

Target	Threshold	Other (Individual)	Others (Collective)
Self	.27***	.40***	.34***
Threshold	-	.28***	.36***

Note. ****p* < .001.

Table S5

Correlations Between Target Judgments, for Moral Behaviors (Study 4b)

Target	Threshold	Other (Individual)	Others (Collective)
Self	.33***	.48***	.51***
Threshold	-	.25***	.38***

Note. ****p* < .001.

Table S6

Correlations Between Targets, for Immoral Behaviors (Study 4b)

Target	Threshold	Other (Individual)	Others (Collective)
Self	.27***	.47***	.40**
Threshold	-	.31***	.31***

Note. ****p* < .001.

Study 4b: Additional Analyses

Per our preregistration, we only made a prediction that self-judgments would again exceed the threshold. Given the social judgments were expected to be influenced by the guilt manipulation, we did not preregister predictions *a priori* for how other and others, under different levels of the guilt manipulation, would compare to the threshold. But we did confirm the preregistered prediction that we would again find that self-positivity would emerge (see below).

As a non-preregistered exploratory analysis, we also tested how the social judgments compared to the threshold under different levels of the guilt manipulation. We found that encouraging people to think that cynicism would lead to more (vs. less) guilt than expected encouraged judgments of others (collectives) to shift from others-negativity to others-neutrality. Specific others were never judged as negative, but always as neutral. We present analyses that detail these results before returning to further consideration of the aspects of the results about which we did not make *a priori* predictions:

We entered the participants' ratings into a mixed model that included Morality (+1 = moral, -1 = immoral), Target (as a categorical variable including *self*, *threshold*, *other*, and *others* as levels), and Guilt (+1 = underestimated, -1 = overestimated) as fixed-effects predictors, as well as all the higher-order interaction terms. Additionally, we treated *participant* and *behavior* as random factors to account for the non-independence of both the participants' multiple judgments as well as the different participants' estimates that concerned the same behavior.

To begin, neither the main effect of guilt manipulation nor its two-way interactions were significant, *F*s < 1.83, *p*s > .176. There was, however, a significant Guilt Manipulation X Morality X Target interaction, *F*(3, 3348.31) = 3.95, *p* = .008. We proceeded to decompose the analyses for each level of the guilt manipulation. When participants were led to believe that people typically overestimate how bad they would feel (guilt manipulation = overestimated), there was a significant Morality X Target interaction, *F*(3, 3348.31) = 45.48, *p* < .001. More specifically, the self was judged significantly more positively than both other and others, *t*s > 5.12, *p*s < .001. Moreover, we again observed self-positivity, *B* = 9.98, *SE* = 1.47, *t*(3348.31) = 6.80, *p* < .001. This is a first confirmation of the preregistered prediction. We observed other-negativity for others, *B* = -9.90, *SE* = 1.74, *t*(3348.31) = -5.70, *p* < .001, but not for a specific other, *B* = 0.39, *SE* = 1.87, *t*(3348.31) = 0.21, *p* = .836. When participants were instead led to believe that people typically underestimate how bad they would feel (guilt manipulation = underestimated), the Morality X Target interaction was also significant, *F*(3, 3348.31) = 25.89, *p* < .001. The self was again judged more positively than both types of other, *t*s > 6.73, *p*s < .001, as well as above the moral threshold, *B* = 9.51, *SE* = 1.47, *t*(3348.31) = 6.48, *p* < .001. This is a second confirmation of the preregistered prediction. This time, however, there was no significant other-negativity for either others, *B* = -3.29, *SE* = 1.83, *t*(3348.31) = -1.80, *p* = .072, or a specific other, *B* = -2.40, *SE* = 1.77, *t*(3348.31) = -1.36, *p* = .174. In other words, it was only when people judged others and were led to think guilt over cynicism was likely to be low that other-negativity emerged. In the other three cases, we observed evidence of other-neutrality.

Some readers may wonder why Study 4's manipulations (i.e., making the target an individual, encouraging people to think they would experience more guilt than expected when being cynical about a collective) eliminated target negativity but did not lead to target positivity. We did not make predictions *a priori* for how the social targets would stack up against the threshold (but only how others and an other might be judged differently as a function of the guilt manipulation) because we did not know how a guilt manipulation that had *all* participants focus and reflect on cynicism—a feature not present in Study 3—might depress social judgment. Although this nuanced question about whether focusing about and writing on cynicism begets cynicism is potentially interesting, its resolution (which we cannot definitively offer) does not change the interpretation of our results that the negativity in judgments of others can be reduced (and even eliminated, if not fully reversed) by manipulating the anticipated negative experience of being cynical.