Target Happiness Attenuates Perceivers' Moral Condemnation of Prejudiced People

Hope Rose[1] [iD], Christopher A. Sanders[2], Chloe Willett[1],
and Laura A. King[1]

Abstract

Five experiments (combined $N = 4,915$) tested the prediction that the moral boost of happiness would persist for social targets with moral failings. In Studies 1 and 2, White and Black participants, respectively, judged happy (versus unhappy) racist targets more morally good. In Study 3, happy (versus unhappy) racist targets were judged more morally good and less (more) likely to engage in racist (good) behavior. Behavioral expectations explained the link between happiness and moral evaluations. Study 4 replicated Studies 1 to 3 in the context of sexism. In Study 5, happy (versus unhappy) targets who engaged in racially biased behavior were evaluated as more morally good, and this effect was explained by behavioral forecasts. Happiness boosts attributions of moral goodness for prejudiced people and does so via expectations for future behavior. Future directions are discussed.

A central concern in getting to know someone is finding out if they are a good person (e.g., Brambilla et al., 2021). How do we know if someone is good? Observing their behavior would likely provide the clearest answer (Reeder & Spores, 1983). Interestingly, not only behavior but also a person's psychological experiences affect moral character inferences. For example, time (Critcher et al., 2013) and effort (Robinson et al., 2017) exerted in thinking about and rendering moral decisions, as well as the emotions that accompany such decisions (Plaks et al., 2022), affect moral attributions to decision makers. One psychological experience that enhances moral attributions is a person's happiness (King & Napa, 1998). In five experiments, we sought to illuminate the nature of this happiness moral boost. To do so, we tested whether happiness would promote moral attributions even in the context of moral failings, specifically, holding racist (Studies 1–2) and sexist (Study 4) attitudes and engaging in racially biased behavior (Study 5). Furthermore, we tested a mechanism of the moral boost of happiness, the expectation that happy people will engage in good behavior (Studies 3–5).

The Happiness Moral Boost

Varied evidence suggests happiness enhances moral attributions. The level of positivity expressed in smiles predicts trustworthiness attributions (Galinsky et al., 2020). Furthermore, smiles lead to higher attributions of trustworthiness as reflected in behavior (in economic games, Centorrino et al., 2015; Krumhuber et al., 2007). Smiling enhances attributions of honesty. For instance, in 37 out of 44 countries, people judged the same targets smiling (versus posing neutral expressions) higher in honesty (Krys et al., 2016). In addition, encounters with smiling targets lead to nonverbal indicators of safety (Miles, 2009). In addition to happy facial expressions, target reports of their subjective happiness affect attributions of global moral goodness (i.e., rating someone as a good person and morally good). In two studies (King & Napa, 1998), participants judged targets reporting themselves to be happy (versus unhappy) more morally good.

Would happiness enhance moral attributions even in the context of moral failings? Answering this question begins to unwrap the processes that might underlie the happiness

[1]University of Missouri, Columbia, USA
[2]Old Dominion University, Norfolk, VA, USA

Corresponding Author:
Hope Rose, Department of Psychological Sciences, University of Missouri, 210 McAlester Hall, Columbia, MO 65211, USA.
Email: hercny@mail.missouri.edu

moral boost. One possibility is that happiness serves as a heuristic (happy = good; Reis et al., 1990). Messages containing incongruent (i.e., positive and negative) information disrupt heuristic processing (Chaiken & Ledgerwood, 2012). Therefore, if the moral boost of happiness represents a heuristic, moral failings should moderate the effect of target happiness, with happiness enhancing moral attributions only (or more strongly) for social targets lacking in such failings.

In contrast, there are reasons to expect the happiness moral boost to persist in the context of moral failings. Genuine smiles have been suggested to serve as costly signals of cooperative interest (Centorrino et al., 2015). Indeed, the level of positivity conveyed in facial expressions facilitates accuracy in judgments distinguishing criminals from noncriminals (Sheldon et al., 2021). Perhaps, happiness plays a substantive role in moral attributions. We hypothesized that happiness would affect attributions of moral goodness, regardless of moral failings, and specifically, that people would judge happy people with moral failings more morally good than unhappy people with those same failings. This hypothesis is rooted in our proposed mechanism for the moral boost of happiness, namely, that happiness implies good behavior.

Explaining the Happiness Moral Boost

Ample research links happiness to good behavior. For exa1mple, compared with unhappy people, happy people recall performing more kind acts in their everyday lives (Krueger et al., 2001; Otake et al., 2006). In turn, meta-analyses demonstrate that agreeableness (being kind, helpful) predicts higher happiness (Wilmot & Ones, 2022). In daily experience, positive mood and prosocial behavior share reciprocal positive relations, with each predicting the other (Snippe et al., 2018). In experiments, prosocial behavior leads to happy feelings (Curry et al., 2018) and induced positive affect leads to prosocial behavior (Lyubomirsky et al., 2005). Summarizing the evidence, scholars have concluded that happiness is best understood as both an outcome and antecedent of prosocial behavior (e.g., Aknin et al., 2018; Hui, 2022).

The effect of happiness on moral attributions might reflect a similar conclusion. If people consider happiness an outcome and antecedent of good actions, then a person's happiness might signal their propensity for such actions. Social perceivers might assume happy people, even those marked by moral failing, must have done or be likely to do something good if they are happy. We hypothesized that the link between happiness and behavioral expectations would explain the effect of target happiness on more global moral attributions.

Prejudice as a Moral Failing

Prejudiced targets provide a promising context for testing our hypotheses for several reasons. Many harmful actions have low base rates, but racist and sexist attitudes are unfortunately commonplace (e.g., Jones & Lloyd, 2021; United Nations News, 2020). In addition, prejudice does not appear to foster distress in prejudiced people (Gilman & Thomas, 2016). Thus, happy racists or sexists may not be utterly disconcerting. At the same time, research links prejudice to especially poor moral character attributions. In two studies (Uhlmann et al., 2014), U.S. participants viewed racially biased behavior (saying a racial slur to oneself; secretly defacing a portrait of Dr. Martin Luther King, Jr.) as more indicative of low moral character than blatantly harmful acts (punching someone; stealing a car). Finally, testing our hypotheses in the context of prejudice allowed us to probe whether people view those high (vs. low) on racist (Studies 1–2) and sexist (Study 4) attitudes as less morally good.

Overview

Five experiments tested the hypotheses that happiness would boost moral evaluations for people with moral failings and that expectations about behavior would explain this moral boost.

In Studies 1 and 2, White and Black participants, respectively, evaluated the moral goodness of targets low/high in racism and happiness. We predicted participants would judge happy (versus unhappy) racist targets as more morally good. Study 3 tested the prediction that expectations for behavior would explain the effect of target happiness on moral judgments of racist targets. Study 4 aimed to conceptually replicate Studies 1 to 3 in the context of sexism. Study 5 probed target prejudicial behavior, testing the predictions that happiness would boost moral evaluations of racist targets who reported engaging in racially biased behavior and that behavioral forecasts would explain this effect.

In Studies 1-2 and 4-5, we did not expect moral failings to moderate the moral boost of happiness (counter to a heuristic explanation). Although we did not preregister these analyses, across these studies, we used Bayes Factors (BF_{01}) to estimate support for the null hypothesis (no interaction) versus the alternative (Faulkenberry, 2021). Interpretation of these values was guided by recommendations from Kinney and colleagues (2021).

Across studies, we tested whether condition effects remained significant controlling for relevant covariates, and tested these variables as moderators. In all studies, we measured conservatism (a positive correlate of both racism, Lin & Alvarez, 2020, and sexism, Prusaczyk & Hodson, 2018). For studies addressing racism, we included Modern Racism (Studies 1, 3, 5), as well as variables relevant to moral views of racism (Tolerance of Racism, White Privilege Remorse, and awareness of the harm of racial prejudice; Studies 3, 5). Study 4 (focusing on sexism) tested participants' gender and sexism.[1] All samples were independent, with previous participants blocked from subsequent studies. Across studies, all

Figure 1. Sample Stimulus, Study 1.
Note. Stimulus for target high in happiness and racism, with happiness items first.

measures, manipulations, and data/participant exclusions are reported in the manuscript and Supplement.

Study 1

White adults judged the moral goodness of targets low/high in racism and happiness. We predicted two main effects, with nonracist (versus racist) targets and happy (versus unhappy) targets judged more morally good. Further, we predicted that happy (versus unhappy) racist targets would be judged more morally good. We tested conservatism and Modern Racism as moderators. Study 1 was preregistered: https://osf.io/sfkdp/?view_only=1c58d97e69fd4224b28557131cd65c00. For measures, data, syntax, and codebook for all studies, see https://osf.io/tkr8c/?view_only=f2c6060c2c72453d919602fa4a4f9b27.

Method

Participants. White U.S.-based adults, $N = 959$ (438 men, 514 women, 4 nonbinary, 3 missing), recruited from Cloud Research, participated for US\$ 0.50. Age ranged from 19 to 84, M (SD) = 43.57 years (13.43). The Supplement, p. 2, provides detailed demographics for all studies.

With no basis for estimating effect size, we sought to recruit ~1,000 participants, based on budget constraints; 1,011 participants enrolled. As preregistered, data were excluded for not completing the study measures (7), not identifying as White (49), and duplicate IP addresses (1; with only the first entry retained). Exclusions were distributed: 18 nonracist/low happiness; 11 nonracist/high happiness; 15 racist/low happiness; and 7 racist/high happiness.

A sensitivity power analysis, with $\alpha = .05$, indicated the sample provided 80% power to detect effects of $d = 0.18$.

Materials and Procedures. Participants saw responses ostensibly provided by a participant in a previous study (after King & Napa, 1998). They were assigned randomly to a 2 (target racism: nonracist vs. racist) $\times$ 2 (target happiness: low vs. high) between-person design. Cell n's = 233, nonracist/low happiness; 239, nonracist/high happiness; 242, racist/low happiness; and 245, racist/high happiness.

Target racism and happiness were manipulated using two items each. For target racism, we created one item, "It is fine for people to have interracial relationships," and used one item from the Modern Racism Scale (McConahay, 1986), "Discrimination against Black people is no longer a problem in America." Nonracist ratings were 6 and 3; racist ratings were 3 and 6. For happiness, the items were, "I am a happy person"; "I often feel unhappy." Low ratings were 2 and 7; high ratings were 6 and 1. The order of racism/happiness items was counterbalanced. Stimuli included 4 distractor items, rated "4" (Figure 1).

With the stimulus visible, participants rated targets on 4 items (adapted from King & Napa, 1998), "morally good," "good," "leads a good life," and "virtuous," from 1 (*not at all*) to 7 (*extremely*), aggregated to measure moral goodness, $\alpha = .92$, M (SD) = 4.25 (1.36). Following exploratory items for another study (see Supplement p. 3), participants rated conservatism, from 1 (*very liberal*) to 7 (*very conservative*), M (SD) = 3.69 (1.82), and 2 items from the Modern Racism Scale, "It is easy to understand the anger of Black people in America" (recoded); "Discrimination against Black people is no longer a problem in America," $r = .77$, M (SD) = 2.73 (1.68).
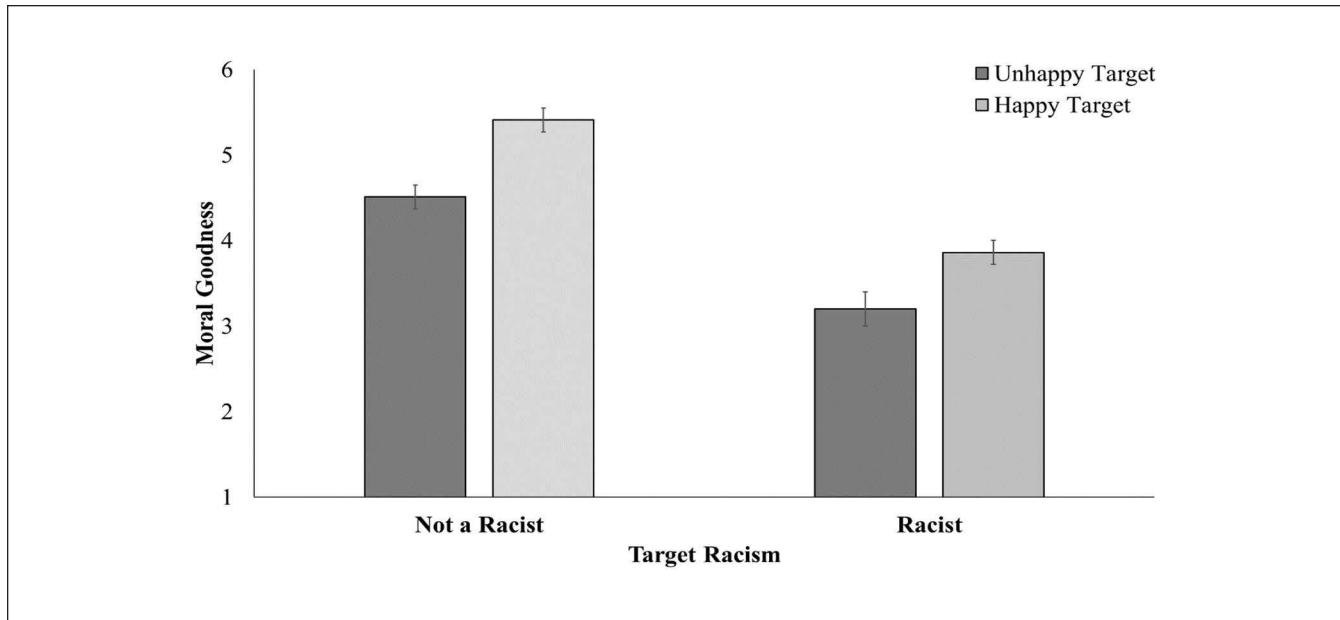
Figure 2. Target Moral Goodness as a Function of Racism and Happiness, Study 1.
Note. N = 959. Error bars are 95% *CIs*. All means differ significantly, *p* < .001.

Manipulations Pilot Test. To avoid drawing attention to racism and happiness, we did not include manipulation checks. Instead, prior to data collection, a pilot test verified that manipulations affected perceptions of target prejudice and happiness without rendering targets atypical. Pilot participants (*N* = 157, White adults from Cloud Research) saw identical manipulations of racism and happiness (see the Supplement pp. 3–5), and rated targets on "prejudiced," "happy," and "similar to the average American" from 1 (*not at all*) to 7 (*extremely*). Participants rated racist targets more prejudiced, *M* (*SD*) = 4.47 (1.87) than nonracist targets, *M* (*SD*) = 2.11 (1.33), *t*(134.99) = 9.08, *p* < .001, *d* = 1.47, and happy targets happier *M* (*SD*) = 5.53 (0.99) than unhappy targets, *M* (*SD)* = 2.28 (1.48), *t*(155) = 16.11, *p* < .001, *d* = 2.57. Target racism did not affect happiness or similarity to the average American (*p*'s > .23). Target happiness did not affect prejudiced ratings, *p* = .21. Similarity to the average American did not differ for unhappy, *M* (*SD*) = 3.71 (1.54) versus happy, *M* (*SD*) = 4.00 (1.25), racist targets, *t*(74) = 0.90, *p* = .37, *d* = 0.21. Thus, manipulations affected perceptions as intended.

Results

Order of racism and happiness items did not significantly affect moral judgments, *p*'s > .153, so analyses collapsed across the order. A 2 (target racism: nonracist versus racist) × 2 (target happiness: low versus high) analysis of variance (ANOVA) on moral goodness revealed the predicted main effects. Participants judged nonracist targets more morally good, *M* (*SD*) = 4.96 (1.06), than racist targets, *M* (*SD*) = 3.55 (1.26), *d* = 1.21, 95% confidence interval [*CI*] = [1.08, 1.35]; *F*(1, 955) = 393.70, *p* < .001. Participants judged happy targets more morally good, *M* (*SD*) = 4.62 (1.24), than unhappy targets, *M* (*SD*) = 3.86 (1.37), *d* = 0.58, 95% *CI* = [0.45, 0.71]; *F*(1, 955) =138.71, *p* < .001. The interaction was not significant, *F*(1, 955) = 3.65, *p* = .056, *d* = 0.01. *BF*$_{01}$ estimates the null hypothesis is 5 times more likely than the alternative, providing moderate support for the null. As Figure 2 shows, as predicted, participants judged happy racists more morally good, *M (SD)* = 3.86 (1.30), than unhappy racists, *M (SD)*= 3.24 (1.14), *d* = 0.51, 95% *CI* = [0.33, 0.69].

Moderation. Conservatism and Modern Racism (*r* = .60, *p* < .001) showed similar moderation patterns. Modern Racism results are in the Supplement (p. 6). A hierarchical regression equation tested for the moderation of condition effects on moral goodness by (mean-centered) conservatism using dummy coded conditions (*nonracist* = 0, *racist* = 1; *unhappy* = 0, *happy* = 1). No interactions involving the happiness condition emerged. The main effects for target racism, β = −.48, target happiness, β = .33, and conservatism, β = −.19 (Step 1, Δ*R*^{2} = .34, *p*'s < .001), were qualified by a significant conservatism × target racism interaction, β = .32 (Step 2, Δ*R*^{2} = .09, *p* < .001). As Figure 3 shows, as conservatism increased, target racism became increasingly irrelevant to moral judgments. Controlling for Modern Racism (β = −.01, *p* = .73), for the conservatism × racism interaction, β = .32 *p* < .001.
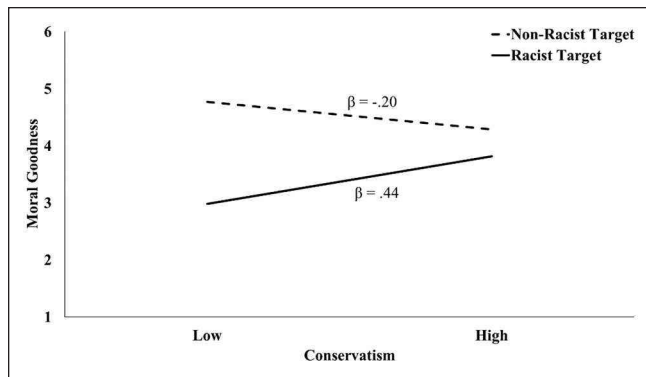
Figure 3. Moral Goodness as a Function of Target Racism X Participant Conservatism, Study 1.
Note. Generated regression lines for participants ±1 *SD* from the mean on conservatism. Slopes are significant, *p* < .001, and differ significantly, *z* = 10.65, *p* < .001.

Brief Discussion, Study 1. As predicted, White participants judged happy racists more morally good than unhappy racists. The absence of an interaction runs counter to a heuristic explanation. Happiness did not raise racists to the level of moral goodness of even unhappy nonracists, but it afforded about a half standard deviation of moral goodness to these targets.

White conservatives (and those endorsing higher Modern Racism) were less likely to consider target racism in moral judgments. Modern Racism results make sense because people endorsing racist views would seem less likely to view such attitudes as morally relevant. Moreover, the manipulation of target racism (inadvertently) contained an item identical to the participant Modern Racism measure.

One explanation for conservatism results is that conservatives did not recognize racist targets as racially prejudiced. Pilot participants rated conservatism as in Study 1, so we addressed this possibility directly in those data. Regressing prejudiced ratings on conservatism (β = .004, *p* = .97) and target racism (β = .61, *p* < .001), the conservatism × target racism was not significant, β = −.17, *p* = .058 (Supplemental Figure S2). Among pilot participants endorsing conservatism ≥ 1 *SD* from the mean, for the effect of target racism on prejudiced ratings, *t*(29) = 2.92, *p* = .007, *d* = 1.08, suggesting White conservatives recognized racist targets as prejudiced. We further probe conservatism as a moderator for target prejudice, in Studies 2 and 4, and racially prejudiced behavior, in Study 5.

White people likely provide a relatively weak test of the happiness moral boost in the context of moral failings because they may not view racist attitudes *as* a moral failing. White (vs. Black) people view racial inequality as less harmful (Horowitz et al., 2019) and endorse higher Tolerance of Racism (e.g., "Someone can be a good person, even if they have some racist views," Hunt et al., 2021). Thus, Study 2 sought to directly replicate Study 1 with Black participants.

Study 2

Black adults rated the moral goodness of targets low/high in racism and happiness. As in Study 1, we predicted nonracist (versus racist) and happy (vs. unhappy) targets would be judged more morally good. Furthermore, we predicted that Black participants would judge happy (versus unhappy) racists as more morally good. We tested conservatism as a moderator. Study 2 was preregistered: https://osf.io/thvrk/?view_only=f86c9f39e4094996a2553021df1ebb5e.

Method

Participants. Black identifying U.S.-based adults, *N* = 918 (404 men, 503 women, 11 nonbinary/an alternative identification), recruited from Cloud Research (*n* = 433) and Prolific Academic (*n* = 485), participated for US$1.00. Age ranged from 18 to 78, *M* (*SD*) = 37.28 years (11.84). As preregistered, data were excluded for not identifying as Black (47) and duplicate IP addresses (57; with only the first entry retained). Exclusions were distributed: 4 nonracist/low happiness; 6 nonracist/high happiness; 14 racist/low happiness; and 14 racist/high happiness.

No main (*p* = .224) or interaction (*p*'s ≥ .305) effects of the data source emerged, so data were merged. Sensitivity power analysis suggested that, with α = .05, this sample provided 80% power to detect an effect of *d* = 0.19.

Procedures and Materials. We randomly assigned participants to the design from Study 1. Because we were (somewhat mistakenly) concerned about recruiting a large Black sample, we preregistered our intent to overrecruit to the cells testing our central prediction (happy vs. unhappy racists). Cell *n*'s = 121, nonracist/low happiness; 119, nonracist/high happiness; 342, racist/low happiness; and 336, racist/high happiness.

We used the stimuli and dependent measure, α = .93, *M* (*SD*) = 3.80 (1.50), from Study 1. Based on Study 1, we removed the item order from the design.

Participants rated conservatism on the item from Study 1, *M (SD)* = 3.18 (1.59), and the 12-Item Social and Economic Conservatism Scale (SECS; Everett, 2013), assessing positive feelings toward 12 political issues (e.g., "Limited Government"), *M (SD)* = 6.44 (1.60), α. = .78.

Results

As in Study 1, participants judged nonracists more morally good, *M* (*SD*) = 5.02 (1.04), than racists, *M* (*SD*) = 3.37 (1.41), *d* = 1.24, 95% *CI* = [1.08, 1.40]; *F*(1, 914) = 290.08, *p* < .001. They judged happy targets more morally good, *M* (*SD*) = 4.14 (1.44) than unhappy targets, *M* (*SD*) = 3.47 (1.50), *d* = 0.46, *CI* = [0.32, 0.59]; *F*(1, 914) = 52.33, *p* < .001. For the target racism X target happiness interaction, *F*(1, 914) = 0.37, *p* = .542, *d* = 0.00. BF_{01} suggests the null
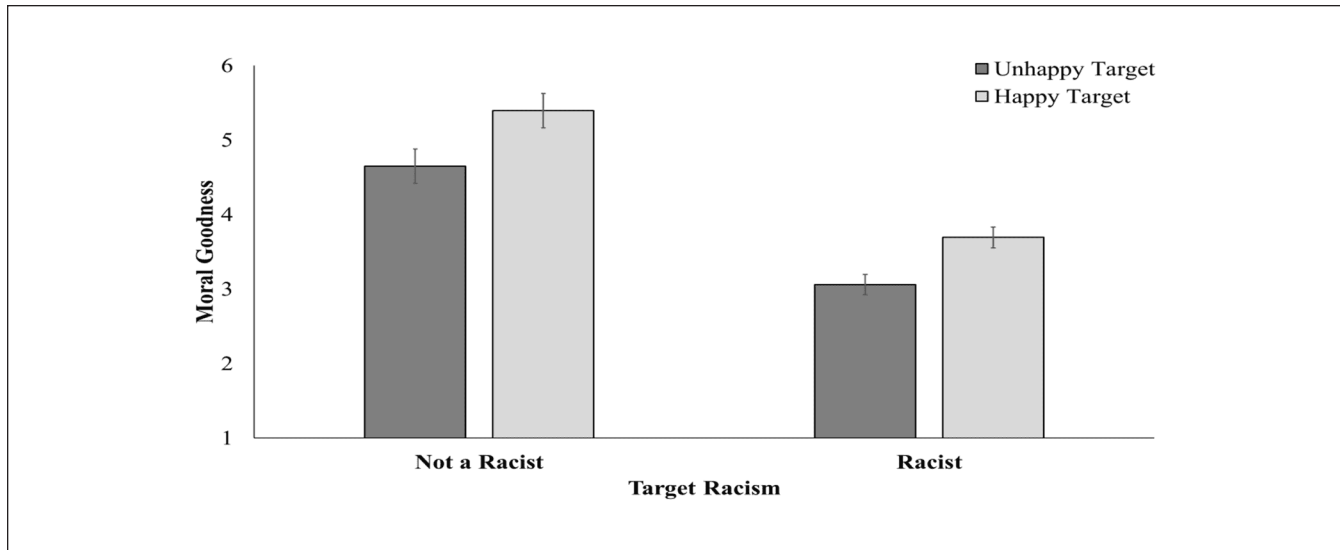
Figure 4. Target Moral Goodness as a Function of Racism and Happiness, Study 2.
Note. N = 918. Error bars are 95% *CI*s. All means differ significantly, *p* < .001.

is 25 times more likely than the alternative, providing strong evidence for the null. As Figure 4 shows, in accord with our preregistered prediction, Black participants judged happy racists more morally good, *M (SD)* = 3.70 (1.34), than unhappy racists, *M (SD)*= 3.06 (1.41), *d* = 0.46, 95% *CI* = [0.31, 0.62].

Moderation. We tested for moderation of condition effects by conservatism, as in Study 1. Because results for the single-item and SCES were similar, we report results for the single item, to enhance comparability to Study 1. For SECS results, see the Supplement (p. 7). No interactions involving target happiness emerged. Significant main effects for target racism, β = −.48, and happiness, β = .23 (Step 1, ΔR^2 = .30, *p*'s < .001), were qualified by a conservatism X target racism interaction, β = .31 (Step 2, ΔR^2 = .02, *p* < .001). As Figure 5 shows, conservatism did not predict moral evaluations for nonracists but positively predicted moral goodness for racist targets.

Brief Discussion, Study 2. Directly replicating Study 1, Black adults judged happy (versus unhappy) racists more morally good. Across studies, White (Study 1) and Black (Study 2) participants did not differ in moral evaluations of nonracist targets, *t*(710) = 0.61, *p* = .55, *d* = 0.05. However, Black (versus White) participants judged racist targets significantly less morally good, *t*(1,110.26) = 2.24, *p* = .013, *d* = 0.13. Despite this mean difference, the *CI*s for the effect size of target happiness for racist targets overlapped considerably, across samples. Thus, Study 2 provides more compelling evidence that happiness boosts moral evaluations in the context of moral failings, among participants viewing racist attitudes as a moral failing.
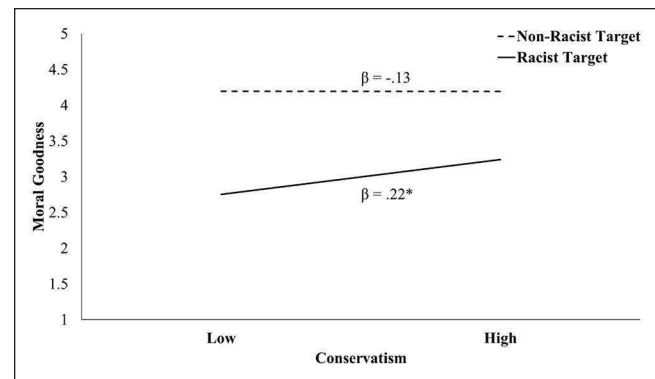

Figure 5. Moral Goodness as a Function of Target Racism X Participant Conservatism, Study 2.
Note. Generated regression lines for participants ±1 *SD* from the mean on conservatism. **p* < .001. Slopes differ significantly, *z* = 5.56, *p* < .001.

Conservatism moderated the effect of target racism on moral evaluations. However, in contrast to Study 1, conservatism only predicted higher evaluations of racist targets. Across studies, the contribution of conservatism to moral goodness for racist targets was significantly stronger for White than Black participants, *z* = 5.87, *p* < .001.

Studies 1 and 2 support our first hypothesis—people judged happy (versus unhappy) racists more morally good. Study 3 tested our second hypothesis that assumptions about behavior explain the happiness moral boost.

Study 3

Participants rated moral goodness and made behavioral forecasts for racist targets low/high in happiness. For behavioral

Table 1. Correlations Among Covariates, Study 3.

Measures	Con.	MR	TOR	WPR	Harm
Conservatism	*.81*				
Modern Racism	.42*	*.77*			
Tolerance of Racism	.37*	.57*	*.80*		
White Privilege Remorse	−.35*	−.38*	−.35*	*.80*	
Awareness of Harm	−.19*	−.54*	−.42*	.26*	*.94*
M(SD)	0.00 (1.0)	3.00(1.65)	3.96(1.17)	3.08(1.71)	5.67(1.49)

Note. N = 1003. *p < .001. Coefficients on the diagonal are α's. Conservatism is a standardized composite.

forecasts, participants rated the likelihood the target would engage in racially biased and more general good/bad behaviors. We predicted participants would judge happy (vs. unhappy) racists more morally good, replicating Studies 1 and 2. We predicted participants would judge happy racists less likely to engage in racially biased behaviors and more likely to engage in good behaviors. We counterbalanced the order of the dependent measures (moral goodness and behavior forecasts) to examine whether behavior forecasts explain moral evaluations or if behavior forecasts are rendered as retrospective justifications of moral attributions (Uhlmann et al., 2014). We predicted that behavioral forecasts would explain the happiness moral boost.

We tested whether condition effects remained significant controlling for conservatism, Modern Racism, Tolerance of Racism, White Privilege Remorse, and awareness of the harm of racial prejudice, and explored these variables as moderators. Study 3 was preregistered https://osf.io/tmnc6/?view_only=a2cc32d340f04719bd8c570d2c436427.

Method

Participants. Participants (N = 1,003; 441 men, 530 women, 8 nonbinary/an alternative identification, 7 prefer not to respond; 76.7% White, 8.9% Asian, 7.5% Black, 7.0% Hispanic/Latino, 1.5% American Indian or Alaska Native, 0.6% Native Hawaiian or Other Pacific Islander, 1.8% Prefer not to respond, 0.2% not listed) were recruited from Cloud Research and received US$0.50. Age ranged from 19 to 85, *M* (*SD*) = 41.97 years (12.77). As preregistered, data were excluded for failing to complete target ratings (19) and duplicate IP addresses (6; with only the first entry retained). Exclusions were distributed: 13 low happiness and 12 high happiness.

Sensitivity power analysis suggested that with α = .05, this sample provided 80% power to detect an effect of *d* = 0.09.

Procedures and Materials. We randomly assigned participants to evaluate happy (*n* = 501) or unhappy (*n* = 502) racist targets, using stimuli (unhappy/happy racist targets) and moral goodness ratings, α = .94, *M* (*SD*) = 3.84 (1.41), from Studies 1 and 2.

Counterbalanced with moral goodness, participants rated the likelihood targets would engage in various behaviors in the future, from 1 (*not likely at all*) to 7 (*extremely likely*). Behaviors included 4 racist (e.g., "make a racist joke"), α = .74, *M* (*SD*) = 4.41 (1.26), 2 bad (e.g., "lie on a resume"), and 2 good (e.g., "donate blood") behaviors. Bad behaviors were reverse scored to create a general good behavior composite, α = .66, *M* (*SD*) = 3.85 (1.02). Bad/good behaviors were judged moderately immoral/moral in pilot testing. The Supplement (pp. 13–19) reports this pilot testing and lists all behaviors used in all studies (including distractor behaviors).

Conservatism measures from Study 2 (single item, *M* [*SD*] = 3.18 [1.59]; SECS, *M* [*SD*] = 58.88 [16.33]) were standardized and aggregated. Two items measured Modern Racism, "It is easy to understand the anger of Black people in America"; and "Black people are getting too demanding in their push for equal rights," (replacing the item from the stimuli).

Participants completed the six-item Tolerance of Racism scale (Hunt et al., 2021; e.g., "Someone can be a good person, even if they have some racist views") and the six-item White Privilege Remorse subscale (from the White Privilege Attitudes Scale; Pinterits et al., 2009, e.g., "I am ashamed of my White privilege"). We created two items tapping awareness of harm, "Racist attitudes are harmful in many ways"; "People's racist attitudes are harmful to themselves and others." Table 1 shows descriptive statistics.

Results

Preliminary Analyses. As Table 1 shows, conservatism, Modern Racism, and Tolerance of Racism are related positively and all related negatively to White Privilege Remorse, and awareness of harm. Remorse and awareness of harm correlated positively. Moral goodness correlated with forecasted racist and good behaviors, *r's* = −.50, .58, respectively, *p*'s < .001. For forecasted racist and good behaviors, *r* = −.60, *p* < .001.

Primary Analyses. A 2 (target happiness: low vs. high) × 2 (order: moral goodness first vs. behaviors first) ANOVA on moral goodness showed the predicted happiness main effect,
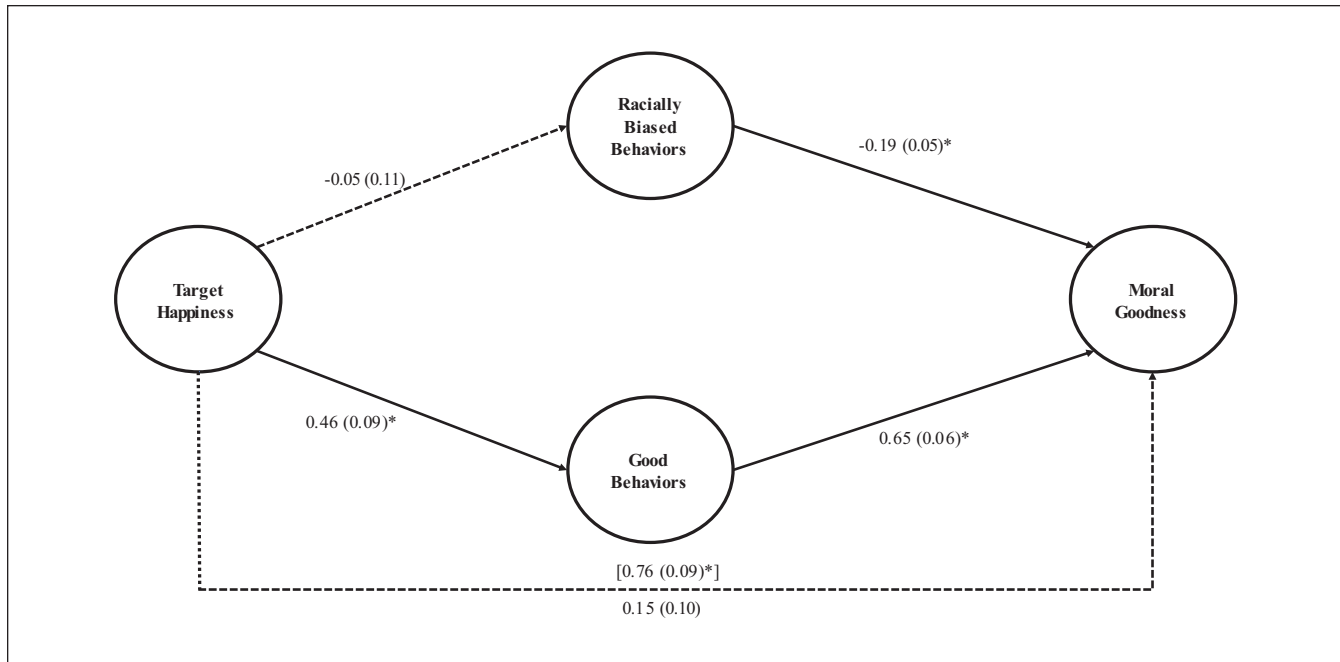
Figure 6. Forecasted Good Behavior Mediates the Effect of Target Happiness on Moral Goodness, Study 3.
Note. n = 489, participants rating behavior first; **p* < .001. Target happiness condition was coded 0 = *unhappy*; 1 = *happy*.

$F(1, 998) = 78.37, p < .001$, with happy racists judged more morally good, $M(SD) = 4.22 (1.42)$ than unhappy racists, $M (SD) = 3.46 (1.29), d = 0.56, CI = [0.43, 0.67]$.[2] Controlling for conservatism, Modern Racism, Tolerance of Racism, White Privilege Remorse, and awareness of harm, the target happiness main effect remained significant, $F(1, 976) = 115.76, p < .001, d = 0.69$.

For behavior forecasts, participants judged happy racists less likely to engage in racially biased behavior, $M (SD) = 4.33 (1.32)$ than unhappy racists, $M (SD) = 4.49 (1.19), d = -0.12 [-0.25, -.002], F(1, 998) = 4.14, p = .042$. They judged happy racists more likely to engage in good behavior, $M (SD) = 4.11 (1.03)$ than unhappy racists, $M (SD) = 3.59 (0.95), d = 0.53 [0.41, .66], F(1, 998) = 71.59, p < .001$. No target happiness × order interactions emerged for behavior forecasts, p's > .18. Controlling for all covariates, the main effects of target happiness remained significant, for racist behaviors, $F(1, 981) = 7.61, p = .006, d = 0.18$; and good behaviors, $F(1, 981) = 84.14, p < .001, d = 0.59$.

Mediation. Similar patterns for moral goodness and behavior forecasts suggest behavior forecasts might explain the happiness moral boost. To probe this possibility, we split the data by order and computed mediational models within each, using PROCESS Macro for SPSS, v22.16.3 (Model 4, Hayes, 2016). As Figure 6 shows, among participants rating behaviors first, behavioral forecasts fully explained the effect of

target happiness on moral goodness. The indirect effect of target happiness through racially biased behavior was not significant, $b (SE) = 0.01 (0.02), 95\% CI = [-0.03, 0.07]$. However, the indirect effect of target happiness through good behavior was significant, $b (SE) = 0.30 (0.06), 95\% CI = [0.18, 0.43]$. For the difference in indirect effects, $z = 4.59, p < .001$. These results support the prediction that the link between happiness and expectations for good (but not racially biased) behavior underlies the moral boost of happiness.

Among participants rendering moral goodness judgments before behavior forecasts ($n = 502$), results suggest a similar conclusion. For comparison purposes, without controlling for moral goodness, target happiness predicted good behavior, $b(SE) = 0.59 (0.09), p < .001$. Controlling for moral goodness judgments, the direct effect of target happiness on good behavior remained significant, $b(SE) = 0.30 (0.07), 95\% CI = [0.16, 0.43], p < .001$. Still, for the indirect effect of target happiness on good behavior through moral goodness, $b(SE) = 0.17 (0.03), 95\% CI = [0.11, 0.24]$.

Moderation. We tested for the moderation of condition effects on the dependent measures. Full results are in the Supplement (pp. 20–25). To summarize, for moral goodness, Modern Racism and Tolerance of Racism more strongly predicted moral attributions for happy (versus unhappy) targets. For racist behaviors, Modern Racism was more positively associated with forecasted racist behaviors for unhappy (vs. happy) targets. For good behavior, conservatism and

Tolerance of Racism were more strongly and positively related to good behavior forecasts for happy (vs. unhappy) targets. White Privilege Shame was unrelated to forecasts for good behavior for unhappy racists but shared a modest negative association with happy racists.

Brief Discussion, Study 3. Replicating Studies 1 to 2, participants judged happy (vs. unhappy) racists more morally good. The effect of target happiness remained significant accounting for relevant covariates. Furthermore, expectations for happy targets' good (but not prejudicial) behavior explained the effect of target happiness on moral goodness attributions. Study 4 tested whether results generalize to sexism.

Study 4

Conceptually replicating Studies 1 to 3, participants rated moral goodness and made behavior forecasts for targets low/high in sexism and happiness. We predicted two main effects, with nonsexist (vs. sexist) targets and happy (vs unhappy) targets judged more morally good. Furthermore, we predicted that happy (vs. unhappy) sexists would be judged more morally good. For behavior forecasts, we predicted participants would judge nonsexist (vs. sexist) and happy (vs. unhappy) targets more likely to engage in good behaviors. We predicted that behavior forecasts would explain the effect of target happiness on moral attributions. We tested participant gender, conservatism, and sexism as covariates and moderators. Finally, although sexism remains an important obstacle for women (United Nations News, 2020), for many, gender inequality is viewed as a problem of the past (Moss-Racusin, 2021). As such, we were interested in probing moral evaluations of targets high (versus low) on sexism. Study 4 was preregistered https://osf.io/dv8t6/?view_only=458cbf431bc4467a91ebe91b7aba33d9.

Method

Participants. Participants ($N = 1,007$, 397 men, 585 women, 13 nonbinary/an alternative identification, 7 missing, 5 prefer not to respond; 77.3% White, 8.7% Asian, 7.9% Black, 7.10% Hispanic/Latino, 3.5% American Indian or Alaska Native, 0.6% Native Hawaiian or Other Pacific Islander, 1.2% prefer not to respond, 0.7% not listed) were recruited from Cloud Research and received US\$0.50. Age ranged from 18 to 94, M (SD) = 42.84 years (12.83). As preregistered, data were excluded for not completing target ratings (34) and duplicate IP addresses (1; with only the first entry retained). Exclusions were distributed: 6 nonsexist/low happiness; 12 nonsexist/high happiness; 6 sexist/low happiness; and 9 sexist/high happiness.

Sensitivity power analysis suggested that, with $\alpha = .05$, this sample provided 80% power to detect an effect of $d = 0.18$.

Procedures and Materials. We randomly assigned participants to a 2 (target sexism: nonsexist vs. sexist) $\times$ 2 (target happiness: low vs. high) between-person design. Cell n's = 254, nonsexist/low happiness; 249, nonsexist /high happiness; 252, sexist/low happiness; 252, sexist/high happiness.

The happiness manipulation and distractors were identical to Studies 1 and 2. To manipulate target sexism, we replaced target racism items with 3 items: "There are many jobs in which men should be given preference over women in being hired or promoted"; "Women should worry less about their rights and more about becoming good wives and mothers" (from the Attitudes Toward Women Scale, Spence & Helmreich, 1972); and "Discrimination against women is no longer a problem in America" (modified from the Modern Racism scale). Nonsexist ratings = 2, 1, 2; sexist ratings = 6, 7, 6.

To broaden the scope of the moral goodness-dependent measure, we added three items to the measure from Studies 1 to 3. These included, "trustworthy," "honest," and "dangerous" (reversed). Because results were the same with and without these items, they were included in the moral goodness composite, $\alpha = .91$, M (SD) = 4.34 (1.19).

Participants rated the likelihood of the target engaging in 4 bad (e.g., "spread a rumor about a friend") and 4 good (e.g., "shovel an elderly neighbor's driveway") behaviors. Reverse-scoring bad behaviors, we created a good behavior composite, $\alpha = .90$, M (SD) = 4.23 (1.12). We intended to counterbalance the order of the dependent measures (moral goodness and good behaviors). However, due to a programming error, counterbalancing was inconsistent, precluding probing order effects.

Conservatism measures from Study 3 (single item, M [SD] = 3.75 [1.76]; SECS, M [SD] = 58.72 [16.41]) were standardized and aggregated, α. = .81. Two items from the Hostile Sexism subscale of the Ambivalent Sexism scale (Glick & Fiske, 1997) measured participant sexism, "Most women interpret innocent remarks or acts as being sexist"; "Women seek to gain power by getting control over men," α = .86, M (SD) = 2.96 (1.57).

Results

Conservatism related positively to sexism, $r = .38$; gender (*coded cisgender men* = 0; *all others* = 1) related negatively to sexism, $-.20$, p's < .001. For moral goodness and behavior forecasts, the correlation was quite strong, $r = .81$, $p < .001$.

For moral goodness, as predicted, participants judged nonsexists more morally good, M (SD) = 4.86 (0.97), than sexists, M (SD) = 3.84 (1.18), $d = 0.98$, 95% CI = [0.81, 1.07]; $F(1, 1003) = 239.82$, $p < .001$. They also judged happy targets more morally good, M (SD) = 4.62 (1.18) than
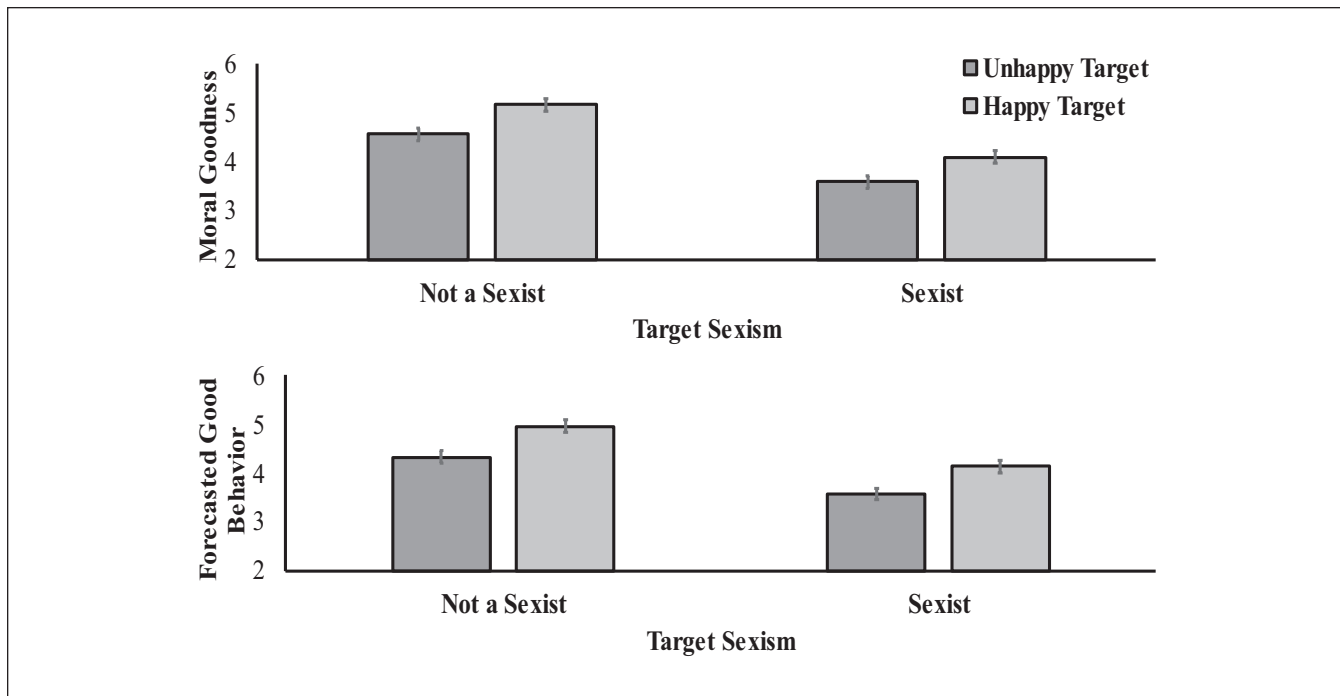
Figure 7. Target Moral Goodness (Top Panel) and Behavior Forecasts (Bottom Panel) as a Function of Target Sexism and Happiness, Study 4.
Note. Error bars are 95% confidence intervals. Within panels, all means differ significantly, $p < .001$.

unhappy targets, M (SD) = 4.07 (1.15), d = 0.53, CI = [0.34, 0.59], $F(1, 1003)$ = 69.84, $p < .001$. The interaction was not significant, $F(1, 1003)$ = 0.65, $p = .420$, $d = 0.06$; BF_{01} = 22, strongly supporting the null. Controlling for participant gender, conservatism, and sexism, main effects for target sexism, $F(1,993)$ = 248.01, d = 1.0, and happiness, $F(1,993)$ = 69.35, d = 0.59, p's < .001, remained significant. As Figure 7 (top panel) shows, as predicted, participants judged happy sexists more morally good, M (SD) = 4.08 (1.16), than unhappy sexists, M (SD)= 3.59 (1.14), d = 0.43, 95% CI = [0.26, 0.61].

Behavior forecasts showed similar results. Participants rated nonsexists more likely to engage in good behaviors, M (SD) = 4.62 (0.98), than sexists, M (SD) = 3.84 (1.10), d = 0.78, 95% CI = [0.62, 0.87], $F(1, 1003)$ = 152.47, $p < .001$. They rated happy targets more likely to engage in good behaviors, M (SD) = 4.53 (1.09) than unhappy targets, M (SD) = 3.94 (1.07), d = 0.50, 95% CI = [0.44, 0.67], $F(1, 1003)$ = 88.06, $p < .001$. The interaction was not significant, $F(1, 1003)$ = 0.22, $p = .64$, $d = 0.00$. BF_{01} = 23, strongly supporting the null. Controlling for all covariates, the main effects remained significant: for target sexism, $F(1,993)$ = 154.73, d = 0.79; for target happiness, $F(1,993)$ = 87.59, d = 0.59, p's < .001. As Figure 7 (bottom panel) shows, as predicted, participants judged happy sexists more likely to engage in good behaviors, M (SD) = 4.13 (1.07), than

unhappy sexists, M (SD)= 3.56 (1.07), d = 0.53, 95% CI = [0.35, 0.71].

Mediation. Mediation models tested the prediction that behavior forecasts explain the effect of target happiness on moral goodness attributions. As Figure 8 (top panel) shows, behavior forecasts fully explain the effect of target happiness on moral goodness. Considering the reverse model (bottom panel), controlling for moral goodness reduced but did not eliminate the significant direct effect of target happiness on behavior forecasts. Both indirect effects were significant and did not differ, z = 1.30, p = .19. For the indirect effect of condition on moral goodness through behaviors, $b(SE)$ = 0.51 (0.06), 95% CI = [0.39, 0.61]. For the indirect effect of condition on behaviors through moral goodness, $b(SE)$ = 0.40 (0.06), 95% CI = [0.30, 0.51].

Finally, target sexism did not moderate the mediation of happiness by behavior forecasts (PROCESS, Model 7), $B(SE)$ = −0.05 (0.12), 95% CI = [−0.25, 0.20], suggesting good behavior explained the effect of happiness on moral goodness regardless of target sexism.

Moderation. We tested for the moderation of condition effects on moral goodness by participant gender, conservatism, and sexism. For gender, significant main effects for target sexism, $F(1, 992)$ = 203.75, and happiness, $F(1, 992)$ =
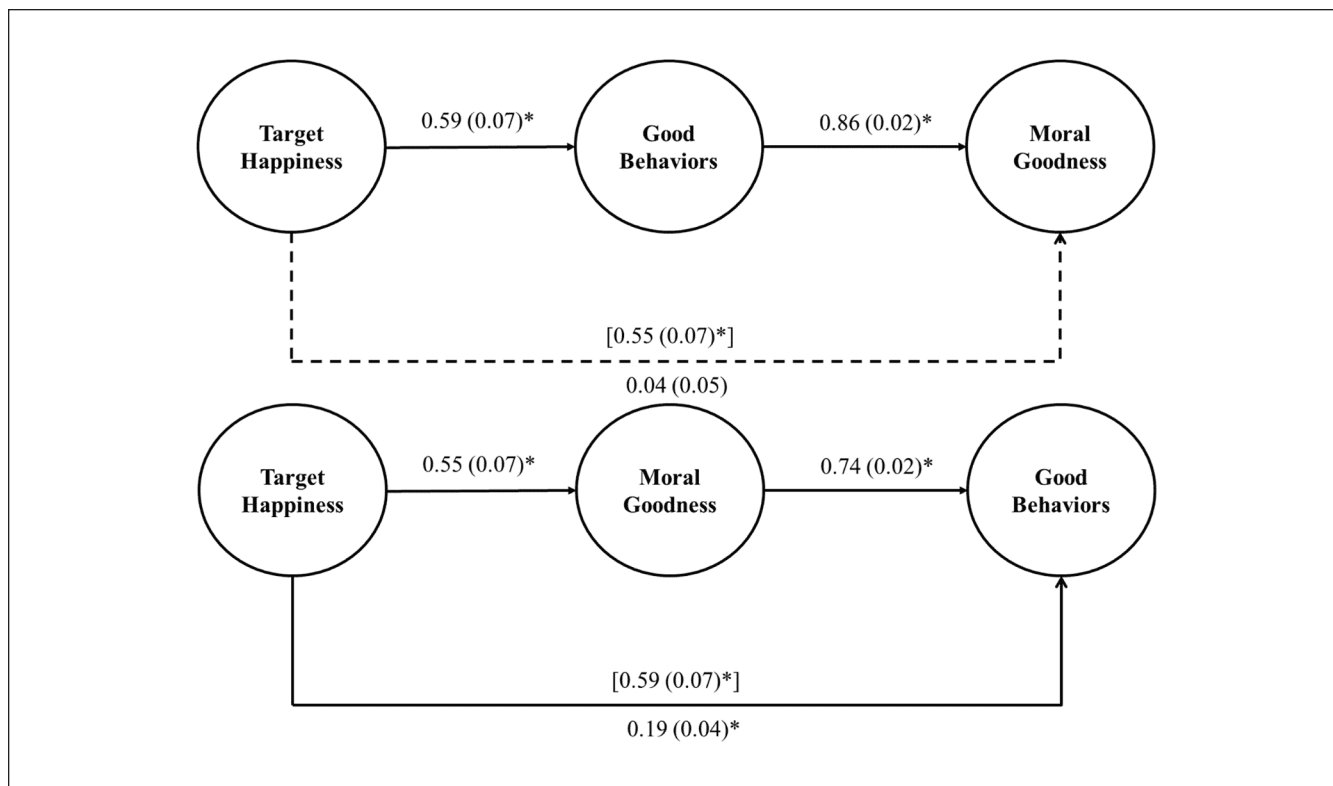
Figure 8. Good Behaviors Mediate the Effect of Target Happiness on Moral Goodness, Study 4.
Note. N = 1007; *p < .001. Target happiness condition was coded 0 = *unhappy*; 1 = *happy*.

64.59, *p's* < .001, were qualified only by the target sexism X participant gender interaction, $F(1,992) = 22.60$ $p < .001$, $d = 0.38$. Women's judgments were more strongly affected by target sexism, for sexists, $M (SD) = 3.69 (1.16)$ versus nonsexists, $M (SD) = 4.97 (0.94)$, $d = 1.21$ 95% CI = [1.04, 1.39]. For men, for sexists, $M (SD) = 4.08 (1.16)$ versus nonsexists, $M (SD) = 4.69 (1.00)$, $d = 0.56$, 95% $CI = [0.36, 0.76]$. Nevertheless, the effect sizes for target happiness across sexist targets were similar and the *CI*s overlapped; for women, $d = 0.45$, 95% $CI = [0.22, 0.67]$; for men, $d = 0.40$, 95% $CI = [0.12, 0.69]$.

Regarding conservatism, no interactions with the happiness condition emerged. The main effects for target sexism, $\beta = -.40$, target happiness, $\beta = .26$, and conservatism, $\beta = -.19$ (Step 1, $\Delta R^2 = .25$, p's < .001), were qualified by a conservatism × target sexism interaction, $\beta = .35$ (Step 2, $\Delta R^2 = .07$, $p < .001$). As Figure 9 shows, as conservatism increased, target sexism became increasingly irrelevant to moral attributions. Controlling for participant sexism ($\beta = .08$, $p = .004$), for the conservatism × target sexism interaction, $\beta = .41$, $p < .001$. Results for participant sexism were similar (see Supplement, p. 26). For behavior forecasts, conservatism and sexism showed similar moderation patterns (Figures S14 and S15).

Brief Discussion, Study 4. Conceptually replicating Studies 1 to 3, happiness boosted moral evaluations of sexist targets,
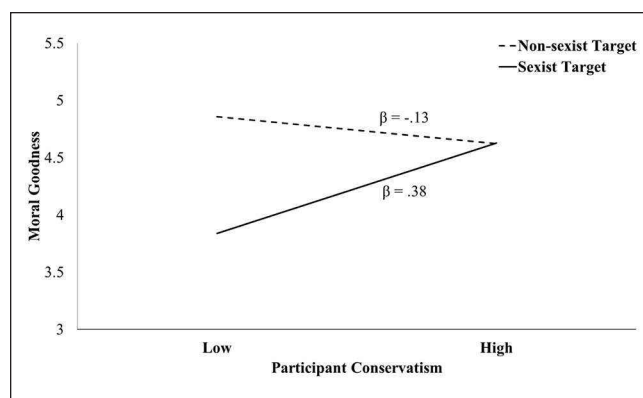

Figure 9. Moral Goodness as a Function of Target Sexism × Participant Conservatism, Study 4.
Note. Generated regression lines for participants ±1 *SD* from the mean on conservatism. Slopes are significant, *p* < .003, for the difference between slopes, *z* = 8.72, *p* < .001.

and the effect of target happiness on moral goodness was explained by behavioral forecasts. Like racists, sexists (versus nonsexists) were judged less morally good. Finally, conservatism moderated the effect of target prejudice on moral goodness, as in Study 1.

Studies 1 to 4 focused on attitudes. Past research suggests that racially biased behavior is especially damaging to moral character inferences (Uhlmann et al., 2014). Study 5 tested

whether happiness would boost moral attributions for targets reporting prejudicial behavior.

Study 5

Participants rated moral goodness and made behavior forecasts for racist targets low/high in happiness who reported engaging in racist behaviors or not. We predicted happy (vs. unhappy) racist targets would be judged more morally good. We predicted that, even among targets reporting racially biased behavior, happiness would boost moral evaluations and behavior forecasts. We predicted behavioral forecasts would explain the moral boost of happiness. We tested the covariates from Study 3. Study 5 was preregistered https:// osf.io/8zkcs/?view_only=26802ad67c6c436abf4eff72eda93 ebc.

Method

Participants. Participants ($N = 1,025$; 370 men, 608 women, 17 nonbinary/genderqueer 6 prefer not to respond; 79.6% White, 8.9% Hispanic/Latino, Asian, 8.6%, 5.7% Black, 2.8% American Indian or Alaska Native, 1.1% Prefer not to respond, 0.5% not listed, 0.4% Native Hawaiian or Other Pacific Islander) were recruited from Cloud Research and received US$0.50. Age ranged from 18 to 80, M (SD) = 42.11 years (13.11). As preregistered, data were excluded for not completing target ratings (29). Exclusions were distributed: 10 no racist behavior/low happiness; 6 no racist behavior/high happiness; 10 racist behavior/low happiness; 2 racist behavior/high happiness.

Sensitivity power analysis suggested that, with $\alpha = .05$, this sample provided 80% power to detect an effect of $d = 0.18$.

Procedures and Materials. We randomly assigned participants to a 2 (racist behaviors: none versus some) $\times$ 2 (target happiness: low vs. high) between-person design. Counterbalanced with the happiness manipulation from Study 3, to manipulate target racist behavior, we showed participants an activities checklist ostensibly completed by the target. We told participants the target had been instructed to check off any activity they had done within the last 7 days on a long checklist of activities, a random subset of which they would see (Figure 10). Distractor items/responses were identical across conditions. Cell n's = 254, no racist behavior/low happiness; 258 no racist behavior/high happiness; 253, racist behavior/low happiness; and 260, racist behavior/high happiness.

Counterbalanced with moral goodness (measured as in Studies 1–3) $\alpha = .94$, M (SD) = 3.37 (1.36), participants made behavioral forecasts for 9 good/bad behaviors (e.g., shovel an elderly neighbor's driveway; spread a rumor about a friend). Bad behavior ratings were reversed, $\alpha = .86$, M (SD) = 3.70 (1.03).

Figure 10. Recent Activities Checklist, Study 5.
Note. Stimulus for target reporting racially biased behavior.

Participant conservatism, M (SD) = 3.66 (1.80) was measured as in Study 1. Modern Racism, $\alpha = .73$, M (SD) = 2.88 (1.51), Tolerance of Racism, $\alpha = .83$, M (SD) = 3.88 (1.24), White Privilege Remorse, $\alpha = .95$, M (SD) = 2.93 (1.59), and harm awareness, $\alpha = .91$, M (SD) = 5.89 (1.32) were measured as in Study 3.

Results

Correlations among covariates were similar to Study 3 (See Supplemental Table S2). Order of the happiness and racist behavior manipulations had no main or interaction effects for either dependent measure (p's < .14). Analyses collapsed across manipulation order. For moral goodness and good behavior forecasts, $r = .68$, $p < .001$.

A 2 (target racist behavior: none vs. some) $\times$ 2 (target happiness: low vs. high) ANOVA on moral goodness revealed the predicted main effects. Participants judged targets reporting no racist behavior more morally good, M (SD) = 3.89 (1.26), than targets reporting racist behavior, M (SD) = 2.85 (1.25), $d = 0.83$, $CI = [0.70, 0.96]$, $F(1, 1019) = 182.60$, $p < .001$. They judged happy targets more morally good, M (SD) = 3.59 (1.41) than unhappy targets, M (SD) = 3.14 (1.26), $d = 0.34$, $CI = [0.22, 0.46]$, $F(1, 1019) = 35.99$, $p < .001$. For the target racist behavior $\times$ happiness interaction, $F(1, 1019) = 0.15$, $p = .70$, $d = 0.00$; $BF_{01} = 29$, strongly supporting the null. Controlling for all covariates, main effects for target racist behavior, $F(1, 997) = 242.39$, $p < .001$, $d = 0.99$, and happiness, $F(1, 997) = 42.88$, $p < .001$, $d = 0.41$, remained significant. As Figure 11 (top panel) shows, as predicted, happy targets who reported racist behaviors were judged more morally good than unhappy targets who did the same, $d = 0.35$, 95% $CI = [0.18, 0.53]$.
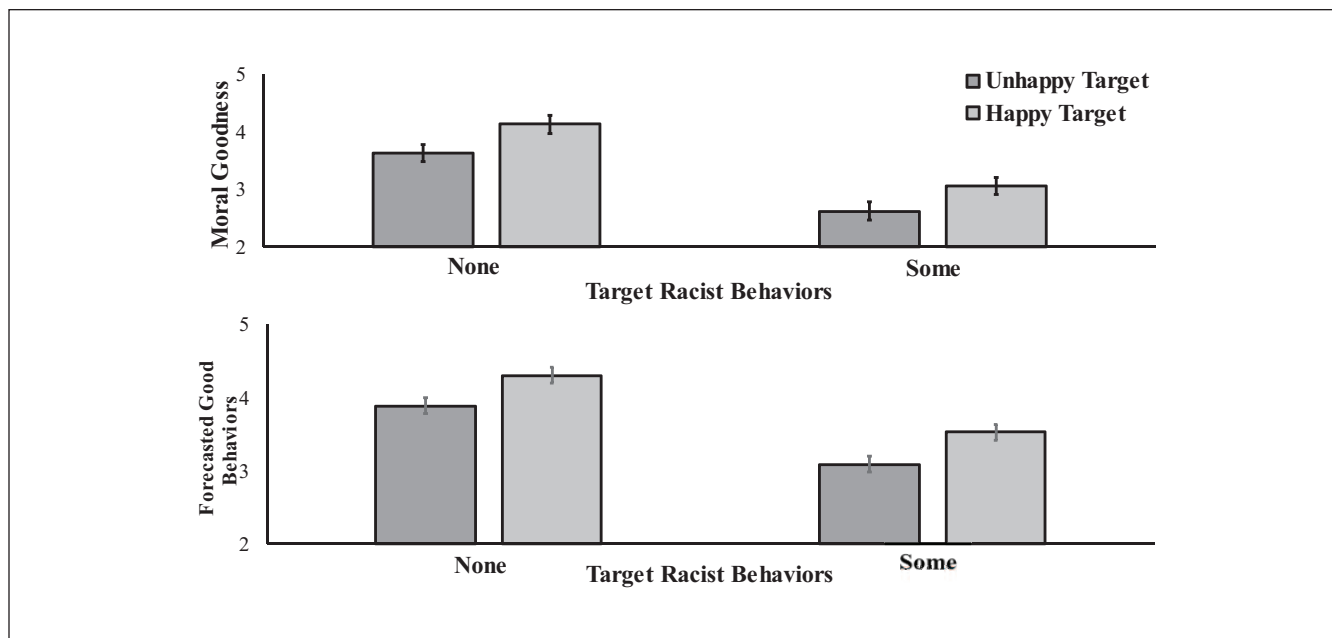
Figure 11. Target Moral Goodness (Top Panel) and Good Behavior (Bottom Panel) as a Function of Target Racist Behavior and Happiness, Study 5.
Note. For moral goodness, $N = 918$; for behaviors, $N = 1,028$. Error bars are 95% CIs. Within panels, all means differ significantly, $p < .001$.

Behavior forecasts showed similar results (Figure 11, bottom panel). Participants judged targets who reported no (versus some) racist behaviors more likely to engage in good behavior, $F(1, 1016) = 192.08$, $p < .001$, $d = 0.87$. They judged happy (versus unhappy) targets more likely to engage in good behavior, $F(1, 1016) = 58.46$, $p < .001$, $d = 0.48$. The interaction was not significant, $F(1, 1016) = 0.07$, $p = .789$, $d = 0.00$; $BF_{01} = 30.90$, strongly supporting the null. Controlling for all covariates, the main effects remained significant: for target racist behaviors, $F(1, 997) = 204.14$, $d = 0.87$, and happiness, $F(1, 997) = 61.28$, $d = 0.48$, $p's < .001$. As expected, happy (vs. unhappy) targets who engaged in racist behaviors were judged more likely to engage in good behaviors, $d = 0.49$, 95% $CI = [0.31, 0.66]$.

Mediation. We split the data by order and computed mediational models within each, as in Study 3. Among participants completing behavior forecasts first, Figure 12 (top panel) shows that behavior forecasts fully explained the effect of target happiness on moral goodness. Among those rendering moral goodness judgments first, Figure 12 (bottom panel) shows moral goodness judgments reduced but did not eliminate the direct effect of target happiness on behavior forecasts. For the indirect effect of target happiness on moral goodness attributions through good behaviors (top model), b $(SE) = 0.50$ (0.08) 95% $CI = [0.34, 0.67]$. For the indirect effect of target happiness on behavior forecasts through moral goodness attributions (bottom model), b $(SE) = 0.19$ (0.06) 95% $CI = [0.07, 0.32]$. For the difference in indirect effects, $z = 2.57$, $p < .001$.

Finally, target racist behavior did not moderate the mediation in these models, either for participants who completed behavior forecasts first, $b(SE) = -0.01$ (0.14), 95% $CI = [-0.32, 0.26]$ or those who completed moral goodness ratings first $b(SE) = 0.01$ (0.12), 95% $CI = [-0.24, 0.23]$, suggesting behavior forecasts explained moral attributions for targets, regardless of racist behaviors (See Supplemental Figures S16 and S17 for mediation within each condition).

Moderation. We tested for the moderation of condition effects on dependent measures. Results were generally similar to Study 3, but there were some (unexpected) three-way interactions (see the Supplement pp 30–36). Notably, the conservatism X target racist *behavior* interaction was not significant, for moral goodness ($\Delta R^2 = .001$; $\beta = .03$, $p = .41$) or behavior forecasts ($\Delta R^2 = .001$; $\beta = .05$, $p = .23$).

Brief Discussion, Study 5. Like the previous studies, for targets reporting racially prejudiced behaviors, happiness boosted moral evaluations, and the effect of target happiness on moral goodness was explained by behavior forecasts. Even in the context of behaviors that past research suggests are indicative of low moral character (Uhlmann et al., 2014), such behavior did not moderate the effect of happiness on moral evaluations. Happy (versus unhappy) people, regardless of their failings, are considered more likely to engage in good behavior and are, therefore, judged more morally good.
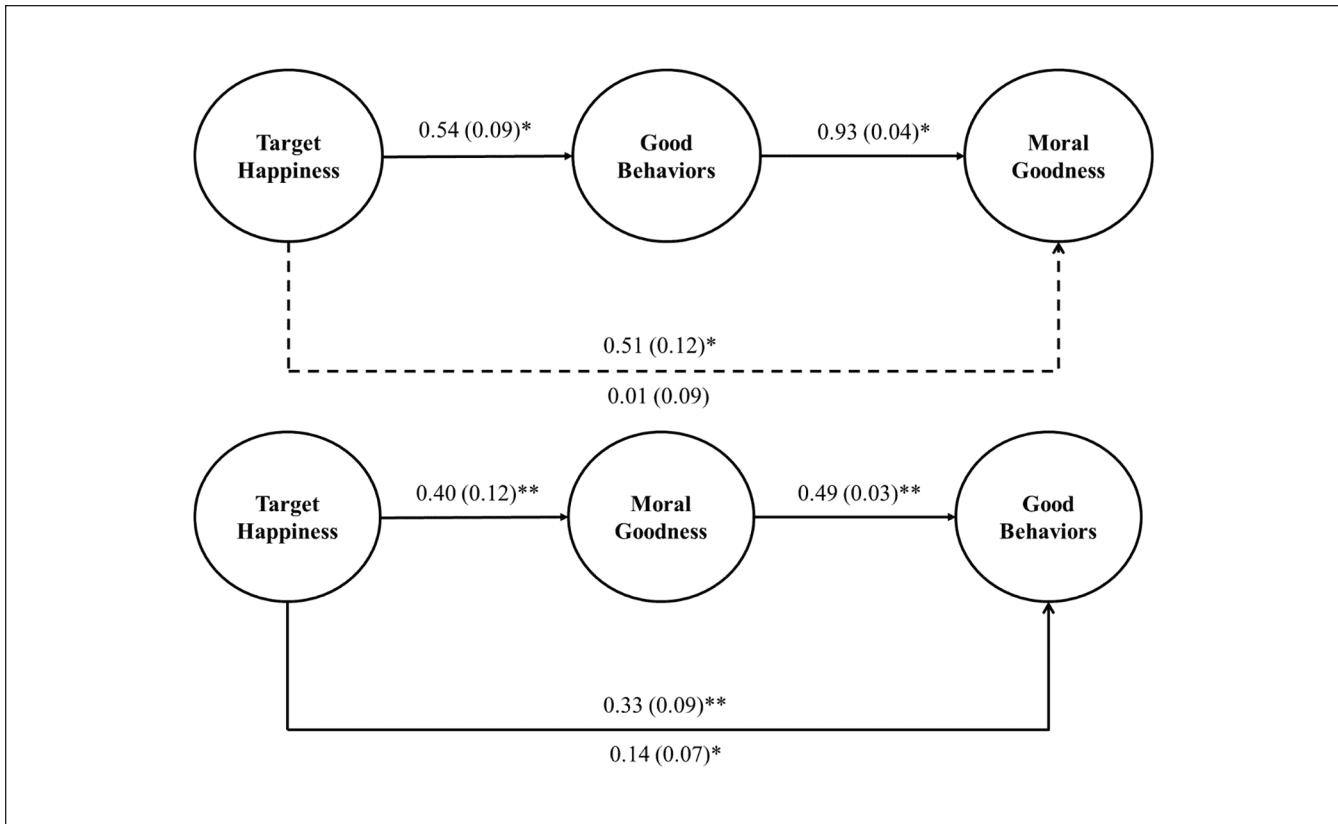
Figure 12. Mediation of Target Happiness on Moral Goodness and Good Behaviors, Study 5.
Note. **p < .001; *p < .05. For the top model, *n* = 520 participants who made behavior forecasts first. For the bottom model, *n* = 496 participants who rated moral goodness first. Target happiness condition was coded 0 = unhappy; 1 = happy.

General Discussion

A person's happiness is one indicator that they are trustworthy, honest, and safe—in other words, a good person. Five studies tested two hypotheses regarding this moral boost of happiness. First, we hypothesized that happiness would boost moral evaluations of people with moral failings. Supporting this hypothesis, amid (large) main effects for target racism (Studies 1 and 2), sexism (Study 4), and racially biased behavior (Study 5), happiness boosted moral attributions, similarly for nonprejudiced and prejudiced targets. As predicted, happy targets with moral failings were judged more morally good than unhappy targets with those same failings. Second, we hypothesized that expectations for good behavior would explain this moral boost. Studies 3 to 5 provide evidence that the link between happiness and behavioral expectations underlies the moral boost of happiness. Global evaluations of moral character are informed by behavioral expectations. In sum, people expect happy people (even those with moral failings) to be more likely to engage in good behavior and such expectations explain the effect of happiness on global moral attributions. These results begin to illuminate the underpinnings of the moral boost of happiness in social perceptions.

Happiness and Morality

The happiness moral boost extends to people with one class of moral failings—holding racist and sexist attitudes and engaging in racially biased behaviors. Across studies, moral failings did not moderate the effects of target happiness on moral attributions, speaking against a heuristic explanation. Furthermore, the happiness moral boost was afforded to happy targets even when their negative attitudes were directed at participants' own groups (Black people in Study 2; women in Study 4). In a sense, these results suggest a moral optimism applied to happy people. Even in the context of moral failings, social perceivers appear to view happiness as an indicator that good behaviors are still likely. Because prejudice is commonplace, people (including targets of prejudice) may be accustomed to separating prejudice from other aspects of a person, including their propensity for nice behavior. In Study 3, racially biased behaviors were less important to such attributions than more mundane good/bad behaviors.

Attributions of moral goodness to happy people may affect views of prejudiced people, generally. People may encounter happy racists or sexists, in everyday life, in the media, or online. Such exposure may promote perceptions

that prejudiced people are good people, despite their harmful attitudes. These perceptions, in turn, may have implications for consequential behavior, including hiring decisions, friendships, and voting.

The role of behavior in the happiness moral boost warrants further investigation. For example, comparing the moral boost provided by prosocial behavior (versus happiness) would provide an important comparative context. Moreover, if expectations of good behavior are a robust mechanism for the happiness moral boost, manipulating target prosocial behaviors should drain happiness of its moral relevance. Future research might push the limits of the happiness moral boost in the context of blatantly harmful moral failings. Happiness might backfire in such instances. Finally, examining whether happiness boosts moral evaluations for targets who possess social identities toward which participants hold prejudices is a promising future research direction.

Considering previously reviewed research suggesting happiness is both a consequence and antecedent of prosocial behavior, expecting happy people to enact good behavior may be a pretty good bet. However, expectations that happiness signals a propensity for good action may be mistaken. Such errors may stigmatize good, unhappy people. Personal happiness emerges from many sources, not all of which are within a person's control (e.g., childhood experiences, Evans & De France, 2022; genetics, Røysamb & Nes, 2019). Information about such factors might mitigate the negative moral connotations of unhappiness. In addition, future research might probe the moral implications of sacrificing personal happiness (i.e., being unhappy) in the service of good action. Errors in attributions of moral goodness to happy people may have negative consequences, as well, for the people making those attributions. Happiness could offer a moral haven to those who engage in harm but are, nevertheless, happy. Ill-intentioned people might exploit the link between happiness and good behavior.

Inferences about moral decision-makers are affected by the feelings that accompany those decisions (Plaks et al., 2022). Similarly, the direct relationship between feelings and moral failings might affect moral evaluations. Future research might test whether a target's emotions (guilt, shame) *about* their failings affect moral attributions.

Conservatism and Moral Views of Prejudiced People

Conservatism moderated the effects of target racism (Studies 1–2) and sexism (Study 4) on moral evaluations, but not target racially biased behavior (Study 5). Why were conservatives less likely to consider prejudicial attitudes in moral evaluations? Conservatism relates negatively to consequentialist thinking (Piazza & Sousa, 2014) so it seems unlikely conservatives were less likely to moralize attitudes. Another explanation may be the link between conservatism and moral

values. Moral Foundations Theory (e.g., Graham et al., 2013) identifies values that contribute to moral judgments, distinguishing between individuating and binding foundations. Unlike liberals, who render moral judgments based primarily on more universal individuating concerns (harm/care, fairness/reciprocity), White U.S. conservatives (Kivikangas et al., 2021) are likely to consider binding factors, including ingroup loyalty, authority, and purity.

Among the binding foundations, ingroup loyalty might be implicated in the present results. Target race and gender were not included in manipulations, but participants may have assumed racist targets were White and sexist targets were men. To the extent that ingroup loyalty implies outgroup animus, conservatives might view prejudice as an extension of ingroup loyalty (and therefore less morally concerning). Future research might address this possibility.

In Studies 3 and 5, when all targets were high in racism, conservatives (and those high in Modern Racism and Tolerance of Racism) evaluated happy (vs. unhappy) targets more positively, suggesting happiness may inspire a benefit of the doubt for those predisposed to excuse racial prejudice.

Political differences in moral evaluations of prejudice may contribute to political polarization. When attitudes take on the character of moral convictions, they are more likely to predict hatred for those with differing views (Skitka et al., 2021). The anonymity of the present studies speaks against the notion that moral attributions were driven by desires to virtue signal. Rather, results suggest for many, racism and sexism are moral concerns.

Limitations

Although these studies feature notable strengths (large samples, direct and conceptual replications across studies), limitations warrant note. To remove visible social identities, manipulations included only target ratings of happiness and prejudice. Future research should probe whether other manipulations (e.g., smiling) produce similar effects. Similarly, research might probe whether behavioral indicators of trust are affected by happiness, even in the context of moral failings. In addition, results may not generalize to biases other than racism and sexism. We probed only explicit prejudices. White Americans hold perpetrators less accountable for discriminatory behavior if their actions emerge from implicit (versus explicit) bias (Daumeyer et al., 2019). Thus, implicit attitudes may be expected to have less moral relevance than explicit attitudes.

Although colorism (Dixon & Telles, 2017) and sexism (United Nations News, 2020) are not limited to the United States, moral judgments of prejudiced people may reflect unique characteristics of the United States. Similarly, although smiling led to higher honesty attributions in China, South Korea, and Japan (Krys et al., 2016), cultural differences might affect the association between happiness and

moral attributions, in the context of moral failings. For example, Japanese (vs. United States) people are less likely to view happiness as an unmitigated good (Uchida, 2010). Cross-cultural evidence of the causal impact of happiness on moral character inferences, in the context of moral failings, is warranted.

Declaration of Conflicting Interests

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

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Hope Rose ⬡ https://orcid.org/0000-0002-8621-4450

Supplemental Material

Supplemental material is available online with this article.

Notes

1. Studies 2 to 5 included religious affiliation and religiosity. Results are in the Supplement, pp. 8–12.
2. For happiness X order, $F(1, 998) = 12.21$, $p < .001$, $d = 0.22$. For unhappy racists, moral goodness was higher if participants rated behaviors before moral goodness, $M (SD) = 3.65 (1.31)$ versus after, $M (SD) = 3.27 (1.25)$, $d = 0.29$. Controlling for covariates, the interaction was not significant, $F(1, 976) = 2.95$, $p = .086$, $d = 0.11$.

References

Aknin, L. B., Van de Vondervoort, J. W., & Hamlin, J. K. (2018). Positive feelings reward and promote prosocial behavior. *Current Opinion in Psychology*, *20*, 55–59.

Brambilla, M., Sacchi, S., Rusconi, P., & Goodwin, G. P. (2021). The primacy of morality in impression development: Theory, research, and future directions. In *Advances in experimental social psychology* (Vol. 64, pp. 187-262). Academic Press.

Centorrino, S., Djemai, E., Hopfensitz, A., Milinski, M., & Seabright, P. (2015). Honest signaling in trust interactions: Smiles rated as genuine induce trust and signal higher earning opportunities. *Evolution and Human Behavior*, *36*, 8–16.

Chaiken, S., & Ledgerwood, A. (2012). A theory of heuristic and systematic information processing. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (Vol. 1, pp. 246–266). Sage.

Critcher, C. R., Inbar, Y., & Pizarro, D. A. (2013). How quick decisions illuminate moral character. *Social Psychological and Personality Science*, *4*(3), 308–315.

Curry, O. S., Rowland, L. A., Van Lissa, C. J., Zlotowitz, S., McAlaney, J., & Whitehouse, H. (2018). Happy to help? A systematic review and meta-analysis of the effects of performing acts of kindness on the well-being of the actor. *Journal of Experimental Social Psychology*, *76*, 320–329.

Daumeyer, N. M., Onyeador, I. N., Brown, X., & Richeson, J. A. (2019). Consequences of attributing discrimination to implicit vs. explicit bias. *Journal of Experimental Social Psychology*, *84*, 103812.

Dixon, A. R., & Telles, E. E. (2017). Skin color and colorism: Global research, concepts, and measurement. *Annual Review of Sociology*, *43*, 405–424.

Evans, G. W., & De France, K. (2022). Childhood poverty and psychological well-being: The mediating role of cumulative risk exposure. *Development and Psychopathology*, *34*(3), 911–921.

Everett, J. A. (2013). The 12-item social and economic conservatism scale (SECS). *PLOS ONE*, *8*(12), Article e82131.

Faulkenberry, T. J. (2021). The Pearson Bayes factor: An analytic formula for computing evidential value from minimal summary statistics. *Biometrical Letters*, *58*, 1–26.

Galinsky, D. F., Erol, E., Atanasova, K., Bohus, M., Krause-Utz, A., & Lis, S. (2020). Do I trust you when you smile? Effects of sex and emotional expression on facial trustworthiness appraisal. *PLOS ONE*, *15*(12), Article e0243230.

Gilman, S. L., & Thomas, J. M. (2016). *Are racists crazy?* New York University Press.

Glick, P., & Fiske, S. T. (1997). Hostile and benevolent sexism. *Psychology of Women Quarterly*, *21*(1), 119–135.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology*, *47*, 55–130.

Hayes, A. F. (2016). Partial, Conditional, and Moderated Moderated Mediation: Quantification, Inference, and Interpretation. Manuscript submitted for publication.

Horowitz, J. M., Brown, A., & Cox, K. (2019, April 9). Race in America, 2. Views of racial inequality. *Pew Research Center*. https://www.pewresearch.org/social-trends/2019/04/09/views-of-racial-inequality/

Hui, B. P. (2022). Prosocial behavior and well-being: Shifting from the "chicken and egg" to positive feedback loop. *Current Opinion in Psychology*, *44*, 231–236.

Hunt, J. S., Folberg, A. M., & Ryan, C. S. (2021). Tolerance of racism: A new construct that predicts failure to recognize and confront racism. *European Journal of Social Psychology*, *51*(3), 551–571.

Jones, J., & Lloyd, C. (2021, July 23). Larger majority says racism against Black people widespread. *Gallup*. https://news.gallup.com/poll/352544/larger-majority-says-racism-against-black-people-widespread.aspx

King, L. A., & Napa, C. K. (1998). What makes a life good? *Journal of Personality and Social Psychology*, *75*(1), 156–165.

Kinney, A. R., Middleton, A., & Graham, J. E. (2021). A Bayesian analysis of non-significant rehabilitation findings: Evaluating the evidence in favor of truly absent treatment effects. *Annals of Physical and Rehabilitation Medicine*, *64*, 101425.

Kivikangas, J. M., Fernández-Castilla, B., Järvelä, S., Ravaja, N., & Lönnqvist, J.-E. (2021). Moral foundations and political orientation: Systematic review and meta-analysis. *Psychological Bulletin*, *147*(1), 55–94.

Krueger, R. F., Hicks, B. M., & McGue, M. (2001). Altruism and antisocial behavior: Independent tendencies, unique personality correlates, distinct etiologies. *Psychological Science*, *12*, 397–402.

Krumhuber, E., Manstead, A. S., Cosker, D., Marshall, D., Rosin, P. L., & Kappas, A. (2007). Facial dynamics as indicators

of trustworthiness and cooperative behavior. *Emotion*, *7*(4), 730–735.

Krys, K., Vauclair, M. C., Capaldi, C. A., Lun, V. M. C., Bond, M. H., Domínguez-Espinosa, A., Torres, C., Lipp, O. V., Manickam, L. S. S., Xing, C., & Antalíková, R. (2016). Be careful where you smile: Culture shapes judgments of intelligence and honesty of smiling individuals. *Journal of Nonverbal Behavior*, *40*, 101–116.

Lin, C., & Alvarez, R. M. (2020). Personality traits are directly associated with anti-Black prejudice in the United States. *PLOS ONE*, *15*(7), Article e0235436.

Lyubomirsky, S., King, L., & Diener, E. (2005). The benefits of frequent positive affect: Does happiness lead to success? *Psychological Bulletin*, *131*(6), 803–855.

McConahay, J. B. (1986). Modern racism, ambivalence, and the modern racism scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91–125). Academic Press.

Miles, L. K. (2009). Who is approachable? *Journal of Experimental Social Psychology*, *45*(1), 262–266.

Moss-Racusin, C. A. (2021). Psychology of gender: Addressing misconceptions and setting goals for the field. *American Psychologist*, *76*(9), 1429–1441.

Otake, K., Shimai, S., Tanaka-Matsumi, J., Otsui, K., & Fredrickson, B. L. (2006). Happy people become happier through kindness: A counting kindnesses intervention. *Journal of Happiness Studies*, *7*(3), 361–375.

Piazza, J., & Sousa, P. (2014). Religiosity, political orientation, and consequentialist moral thinking. *Social Psychological and Personality Science*, *5*(3), 334–342.

Pinterits, E. J., Poteat, V. P., & Spanierman, L. B. (2009). The White Privilege Attitudes Scale: Development and initial validation. *Journal of Counseling Psychology*, *56*(3), 417–429.

Plaks, J. E., Robinson, J. S., & Forbes, R. (2022). Anger and sadness as moral signals. *Social Psychological and Personality Science*, *13*(2), 362–371.

Prusaczyk, E., & Hodson, G. (2018). Left-right differences in abortion policy support in America: Clarifying the role of sex and sexism in a nationally representative 2016 sample. *Personality and Individual Differences*, *127*, 22–25.

Røysamb, E., & Nes, R. B. (2019). The role of genetics in subjective well-being. *Nature Human Behaviour*, *3*, 3.

Reeder, G. D., & Spores, J. M. (1983). The attribution of morality. *Journal of Personality and Social Psychology*, *44*(4), 736–745.

Reis, H. T., Wilson, I. M., Monestere, C., Bernstein, S., Clark, K., Seidl, E., Franco, M., Gioioso, E., Freeman, L., & Radoane, K. (1990). What is smiling is beautiful and good. *European Journal of Social Psychology*, *20*(3), 259–267.

Robinson, J. S., Page-Gould, E., & Plaks, J. E. (2017). I appreciate your effort: Asymmetric effects of actors' exertion on observers' consequentialist versus deontological judgments. *Journal of Experimental Social Psychology*, *73*, 50–64.

Sheldon, K. M., Corcoran, M., & Trent, J. (2021). The face of crime: Apparent happiness differentiates criminal and non-criminal photos. *The Journal of Positive Psychology*, *16*(4), 551–560.

Skitka, L. J., Hanson, B. E., Morgan, G. S., & Wisneski, D. C. (2021). The psychology of moral conviction. *Annual Review of Psychology*, *72*, 347–366.

Snippe, E., Jeronimus, B. F., aan het Rot, M., Bos, E. H., de Jonge, P., & Wichers, M. (2018). The reciprocity of prosocial behavior and positive affect in daily life. *Journal of Personality*, *86*(2), 139–146.

Spence, J. T., & Helmreich, R. (1972). Who likes competent women? Competence, sex-role congruence of interests, and subjects' attitudes toward women as determinants of interpersonal attraction1. *Journal of Applied Social Psychology*, *2*(3), 197–213. https://doi.org/10.1111/j.1559-1816.1972.tb01272.x

Uchida, Y. (2010). A holistic view of happiness: Belief in the negative side of happiness is more prevalent in Japan than in the United States. *Psychologia*, *53*(4), 236–245.

Uhlmann, E. L., Zhu, L., & Diermeier, D. (2014). When actions speak volumes: The role of inferences about moral character in outrage over racial bigotry. *European Journal of Social Psychology*, *44*(1), 23–29.

United Nations News (2020, March 5). *Report reveals nearly 90 per cent of all people have "a deeply ingrained bias" against women*. https://news.un.org/en/story/2020/03/1058731

Wilmot, M. P., & Ones, D. S. (2022). Agreeableness and its consequences: A quantitative review of meta-analytic findings. *Personality and Social Psychology Review*, *26*(3), 242–280.