What's In It for Me? Beliefs About Relative Costs to Well-Being Explain Why People Deprioritize Moral Improvements

Jessie Sun[1] & Jonathan Z. Berman[2]

[1]Washington University in St. Louis
[2]London Business School

Author Note

All study materials and the data and code required to reproduce the results reported in this manuscript are available at https://osf.io/xq6sf. Correspondence concerning this article should be addressed to Jessie Sun, Department of Psychological and Brain Sciences, Washington University in St. Louis, 1 Brookings Drive, St. Louis, MO 63130. Email: jessie.sun@wustl.edu.

Acknowledgments

We thank David Condon and Mijke Rhemtulla for their advice on planned missing data designs, and Sydney Scott for providing feedback on a draft of this manuscript.

Statements and Declarations

Ethical considerations: Data collection procedures were approved by the Institutional Review Board at the University of Pennsylvania (IRB ID: 831767), the Research Ethics Committee at London Business School (REC ID: REC771-11102024), and the Institutional Review Board at Washington University in St. Louis (IRB ID: 202309103).

Consent to participate: Participants provided informed consent via the Qualtrics survey.

Consent for publication: Not applicable

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

Funding statement: The authors did not receive funding for this work.

Data availability: Data and analysis code are available at https://osf.io/xq6sf.

Word count: 10,031

Version: Manuscript accepted for publication at *Personality and Social Psychological Bulletin* (July 24, 2025).

Abstract

Most people are interested in improving themselves, but they show less interest in improving on moral traits. Why don't people particularly want to be more moral, and why do people prioritize improving certain traits more than others? Across four preregistered studies of U.K.-based CloudResearch ($N_{\text{Study 1}}$ = 252; $N_{\text{Supplemental}}$ = 110) and Behavioral Research Lab ($N_{\text{Study 2a}}$ = 303, $N_{\text{Study 2b}}$ = 301) participants, we test four classes of explanations. Results rule out explanations based on the ideas that moral traits are seen as more difficult to change and as either more or less causally central. Instead, people are less interested in moral improvements because they believe (a) that they are already highly moral and (b) that nonmoral improvements would more effectively improve their happiness and goal attainment. These results clarify the perceived tradeoffs between well-being and morality and show that personal well-being is a central motivation for personality change.

Keywords: morality, well-being, personality change, self-improvement, goals

What's In It for Me? Beliefs About Relative Costs to Well-Being Explain Why People Deprioritize Moral Improvements

All around the world (Baranski et al., 2021) and across the lifespan (Hudson & Fraley, 2016), most people are interested in improving themselves (Hudson & Roberts, 2014). However, people are much less interested in improving more morally relevant traits (e.g., compassion, honesty, loyalty) compared with less morally relevant traits (e.g., sociability, organization, anxiety; Sun & Goodwin, 2020; Thielmann & de Vries, 2021). If being moral is so highly valued both from a personal (Aquino & Reed, 2002; Monin & Jordan, 2009) and social perspective (Goodwin et al., 2014; Hartley et al., 2016), then why aren't people particularly interested in becoming more moral? And why *do* people particularly want to change certain aspects of themselves? Here, we test four classes of explanations that potentially explain why people want to change some traits more than others, and why people deprioritize moral traits more specifically (see Table 1 for an overview).

The Deficit-Reduction Hypothesis

The Deficit-Reduction Hypothesis argues that people want to improve traits that they perceive greater deficiencies on. Past work suggests that people tend to perceive having higher levels of moral (vs. nonmoral) traits (Sun & Goodwin, 2020), perhaps due to especially strong self-serving biases in the moral domain (Epley & Dunning 2000; Tappin & McKay, 2017). Therefore, perhaps the most obvious explanation for why people are less interested in improving moral traits is that they see less room for improvement on such traits. Yet, research shows that, controlling for current levels, people are still less interested in moral improvements (Sun & Goodwin, 2020). In other words, while perceived deficiencies partially explain why people deprioritize moral improvements, additional explanations are needed.

Table 1

Overview and Summary of Support for Various Hypotheses About Why People Are Less Interested in Moral Improvements

	Relation with Change Goals		Relation with Morality	
Hypothesis	Prediction	Supported?	Prediction	Supported?
1. Deficit reduction				
Low levels	People are more interested in improving traits that they have lower levels of.	✅ (all studies, but not when controlling for suboptimality)	People report having higher levels of moral (vs. nonmoral) traits.	❌ (S1) ✅ (S2a, S2b, SS)
Suboptimal levels	People are more interested in improving traits for which their actual levels fall shorter of their optimal levels.	✅ (S2b, SS)	People report having closer-to-optimal levels of moral (vs. nonmoral) traits.	✅ (S2b, SS)
2. Causal centrality				
Identity continuity	People are less interested in improving traits that are more causally central.	🔄 (S1)	Moral traits are perceived as being more causally central.	❌ (S1)
Personality enabling	People are more interested in improving traits that are more causally central.	✅ (S1, but only in SPM)	Moral traits are perceived as being less causally central.	❌ (S1)
3. Personal fulfillment	People are more interested in improving traits that would better improve their well-being.	✅ (all studies)		
Absolute costs			Moral improvements are perceived to reduce well-being.	🔄 (all studies)
Relative costs			Nonmoral (vs. moral) improvements are perceived to improve well-being to a greater extent.	✅ (all studies; only happiness and goal attainment)
4. Malleability				
Difficulty	People are less interested in improving traits that seem more difficult to change.	🔄 (SS)	Moral traits are perceived as more difficult to change.	🔄 (SS)
Controllability	People are more interested in improving traits for which changes are perceived to be more within their control.	🔄 (SS)	Moral trait changes are perceived as being less within one's control.	🔄 (SS)

Note. For each hypothesis to explain why people are less interested in moral improvements, both predictions (relation with change goals and relation with morality) must be supported. Hypotheses for which both predictions are generally supported are shown in **boldface.** ✅ = prediction supported, ❌ = prediction not supported, 🔄 = evidence opposite to the prediction. S1 = Study 1, S2a = Study 2a, S2b = Study 2b, SS = Supplemental Study. The malleability hypothesis was only tested in the Supplemental Study (see Supplemental Material, Section 5). The suboptimal levels version of the deficit reduction hypothesis was only tested in Study 2b and the Supplemental Study. The personality enabling and identity continuity hypotheses were only tested in Study 1. Study 1 tested a narrow version of the personal fulfillment hypothesis (focusing on happiness); the other studies tested a more expansive version (capturing a broader range of well-being outcomes). Here, we use "moral" and "nonmoral" as short-hand for "more morally relevant" and "less morally relevant." SPM = single-predictor model.

An additional possibility that past research has not yet addressed is that people think they already have close to optimal levels of morality, where the optimal (i.e., best, most adaptive, and well-suited) level may not be the highest possible level. In other words, perceived deficiencies might better be captured by considering distance from one's optimal levels of the trait, rather than distance from the highest possible level of the trait. Here, we examine the possibility that people view themselves as being close to optimally moral and test whether people still deprioritize moral traits after controlling for perceived optimality.

The Identity-Continuity and Personality-Enabling Hypotheses

A second class of explanations suggests two competing hypotheses about the role of causal centrality. On the one hand, people might be averse to the idea of changing causally central traits because such changes would fundamentally change who they are as a person (Chen et al., 2016). We refer to this as the Identity-Continuity Hypothesis. Alternatively, people might seek to improve traits that are most causally central, because these traits give them greater leverage for improving their other personality traits more broadly. For instance, a person may believe that becoming less anxious would also enable them to become more sociable, compassionate, and efficient. If so, people may prioritize improving these more causally central traits because such improvements would give them the greatest "bang-for-buck" in terms of having a more expansive positive impact on their personalities. We refer to this as the Personality-Enabling Hypothesis.

Past work has suggested that moral (vs. nonmoral) traits are more central to personal identity (De Freitas et al., 2018; Strohminger et al., 2017). This might imply that people believe that moral improvements would have greater positive impacts on their personalities (and therefore, that the Personality-Enabling Hypothesis cannot explain why people are *less* interested

in being more moral). However, this work has primarily equated the centrality of morality with notions of personal identity and the "true self" (e.g., Bench et al., 2015; Molouki & Bartels, 2017; Strohminger & Nichols, 2015). This is distinct from the question of which traits are deemed to have larger positive impacts on one's other traits.

To our knowledge, only one study to date has addressed this latter question. For each of 16 features of personal identity, Chen and colleagues (2016) asked participants to report which and how much of the other 15 features was caused by the target feature. They found that people believe that moral traits are more causally central (i.e., exert causal effects on other aspects of personal identity) than nonmoral personality traits. However, because Chen and colleagues examined a wide range of aspects of personal identity, they sampled only three moral and three nonmoral personality traits. Other nonmoral traits (e.g., depression, energy level) may well have been more causally central had they been included. Thus, it is an open question as to whether people believe that nonmoral improvements would provide greater leverage than would moral improvements, and whether this is why they prioritize nonmoral improvements. Moreover, given that people believe that one's capacity to inflict either goodness or harm onto the world is conditional on being agentic, competent, or strong-willed (Baumeister & Exline, 1999; Gai & Bhattacharjee, 2022; Gray & Wegner, 2009; Hofmann et al., 2018; Landy et al., 2016), it is at least plausible that people might believe that improving certain nonmoral traits would improve moral traits.

The Personal Fulfillment Hypothesis

A third class of explanations argues that people prioritize changing traits that would facilitate improvements to their personal well-being. We refer to this as the *Personal Fulfillment Hypothesis*. In line with this hypothesis, Gander and Wagner (2024) found that people were more

interested in improving character strengths (positively valued traits described by the VIA classification; Peterson & Seligman, 2004) that: a) were more positively associated with well-being; and b) they believed contributed more positively to their own well-being. Although morality is a deeply felt value, it is unclear how people prioritize being moral relative to other values such as being happy, living a meaningful life, or having strong social connections. Even if people do want to be more moral, they may be more interested in making changes that would maximize their well-being. Here, we test whether people are less interested in moral improvements because they intuit a tradeoff between well-being and morality.

For one, people might be less interested in improving moral traits because they believe that such changes would have negative consequences for their well-being in an *absolute* sense. That is, they may believe that becoming more moral would reduce well-being. In line with this, people often donate more to charity when doing so involves enduring pain or exerting effort (vs. being enjoyable or easy; Olivola & Shaif, 2013). When judging others, people praise those who have sacrificed in service of acting morally (Bigman & Tamir, 2016; Johnson, 2022; Erlandsson et al., 2020). And, those who make moral proclamations are sometimes deemed to be hypocrites unless they have sufficiently sacrificed (Effron & Miller, 2015; O'Connor et al., 2020). Together, this suggests that people associate morality with sacrifice and may believe that being more moral requires absolute sacrifices to one's own well-being.

Alternatively, even if people do not think that improving their morality would come at an absolute cost to their well-being, they might believe that nonmoral improvements would lead to *relatively* greater well-being benefits compared with moral improvements. Whereas some studies suggest an association between morality and sacrifice, other studies suggest that people intuit well-being benefits of moral traits and moral improvement. For example, people believe that all

24 character strengths (positively valued traits described by the VIA classification; Peterson & Seligman, 2004)—which include morally relevant traits such as honesty, kindness, and fairness—are good for a person's own happiness and well-being (Gander & Wagner, 2024). When people think of a behavior or trait they want to change about themselves to become more moral, they also tend to believe that making that change would improve their own well-being (Sun et al., 2024). However, neither of these studies test whether people expect similar or smaller well-being benefits of more vs. less morally relevant improvements.

In sum, the Personal Fulfillment Hypothesis might explain why people are less interested in improving their moral traits if people believe that (a) moral improvements would reduce their well-being in an *absolute* sense, or (b) nonmoral improvements would lead to *relatively* greater improvements to their well-being than would moral improvements.

The Malleability Hypothesis

Finally, perhaps people consider the feasibility of changing various traits and believe that moral (vs. nonmoral) traits are less malleable. However, past work shows that people tend to rate moral traits as being *more* changeable (Sun & Goodwin, 2020). Accordingly, we did not include this explanation at the outset of the project. However, for comprehensiveness (and in response to feedback), we examined two alternative operationalizations of malleability in a Supplemental Study (see Supplemental Material, Section 5). The Supplemental Study shows that (a) people believe that making moral improvements would be *less* difficult and *more* within one's control, and that (b) people are *more* interested in improving traits for which changes are perceived to be *more* difficult and *less* within one's control. Both findings are the opposite of what would be predicted by the Malleability Hypothesis. For concision, we therefore focus on the other three classes of explanations in the main text.

The Present Research

We investigate why people are more interested in improving certain traits within themselves, and more specifically, why they deprioritize more morally relevant improvements. Across three preregistered studies, we test the roles of perceived deficiencies (the Deficit-Reduction Hypothesis; Studies 1–2b), expected impacts on one's other traits (the Personality-Enabling and Identity-Continuity Hypotheses, respectively; Study 1), and expected consequences for one's well-being (the Personal Fulfillment Hypothesis; Studies 1–2b). In a Supplemental Study (see Supplemental Material, Section 5), we also test and rule out the Malleability Hypothesis.

Ethics and Open Practices Statement

Data collection procedures were approved by the IRB at the University of Pennsylvania (IRB ID: 831767), the REC at London Business School (REC ID: REC771-11102024), and the IRB at Washington University in St. Louis (IRB ID: 202309103). We preregistered stopping rules, exclusions, and analysis plans at https://osf.io/hx8fg (Study 1), https://osf.io/8cnuq (Study 2a), and https://osf.io/pfazj (Study 2b). All study materials and the data and analysis code are available at https://osf.io/xq6sf. We report all data exclusions, all manipulations, and all measures in all studies. All preregistered analyses have been reported across the main text and supplemental materials, and any deviations from the preregistered plan are clearly marked.

Study 1

Study 1 provides initial tests of the Personality-Enabling, Identity-Continuity, and Personal Fulfillment Hypotheses. To assess the former two, participants imagine how improving each of their personality traits would affect their other traits. To assess the Personal Fulfillment Hypothesis, participants imagine how various trait improvements would affect their happiness.

We preregistered research questions rather than directional hypotheses. However, it is worth noting that we introduced the competing Identity-Continuity Hypothesis in response to feedback (rather than at the outset of the project).

Method

Participants and Procedure

CloudResearch participants were paid $1.25 for an online survey that took approximately 8–10 minutes to complete. For this initial study, we focused on a manageable set of 10 traits (for detailed justifications, see Supplemental Material, Section 1.1), of which three were moral (morality, honesty, compassion) and seven were nonmoral (sociability, energy level, self-control, anxiety, aesthetic sensitivity, intelligence, happiness). In subsequent studies, we expand this to 22 input traits. Participants viewed the 10 traits' definitions (see Appendix A), before completing a comprehension check in which they were required to correctly match up each trait with its definition. Those who failed the comprehension check were automatically screened out of the survey. Those who passed completed a random subset of items about their beliefs about how improving one trait would change each of the other 9 traits (described below). Finally, participants reported their current trait levels and change goals (described below).

We aimed to recruit 250 participants on the basis that this would provide an average of 150 responses for each change belief item (given our planned missing data design, described below). 281 surveys were submitted. As preregistered, we excluded participants who failed the comprehension check. The final sample comprised 252 participants (134 women, 114 men, 3 preferred not to say) who ranged in age from 20 to 77 years (M = 39.57, SD = 12.36) and identified as White (n = 197), Black/African American (n = 24), Asian (n = 12), Other or

Multiple (n = 10), Hispanic/Latino (n = 7), Pacific Islander (n = 1), or did not report their race (n = 1).

Measures

Change Goals. Following Thielmann and de Vries (2021), we measured change goals using one item per trait: "To what extent would you like to change how [trait] you are:" (–3 = *I want to be much less [trait] than I currently am*, 0 = *I don't want to change how [trait] I am*, +3 = *I want to be much more [trait] than I currently am*).

Perceived Deficiencies. We used 2 to 3 items to measure participants' self-reported current levels of each trait. The items and their sources are in Appendix A. Each item was preceded with the stem, "I am someone who…". Participants indicated their agreement with each item on a 5-point scale anchored by 1 (*strongly disagree*) and 5 (*strongly agree*). After computing the trait composites, we reversed the scores for all traits except anxiety, such that higher scores represented greater perceived deficiencies on a given trait.

Beliefs about the Consequences of a Trait Change. Participants were instructed to "Please do your best to imagine what you would be like if you changed one aspect of yourself enough that it makes a noticeable difference in your daily life." They then reported their beliefs about how improving one "input" trait (e.g., "If I became more [moral], I would:") would change another "outcome" trait (e.g., sociability). Responses were made on a 7-point scale anchored with –3 (*Become much less [sociable] than I currently am*), 0 (*Remain as [sociable] as I currently am*), and +3 (*Become much more [sociable] than I currently am*). For anxiety, a reduction in the trait reflects an improvement. Accordingly, the input trait item was phrased in terms of becoming "less anxious," and items that involved the outcome trait of anxiety were reversed, such that higher scores capture reductions in anxiety.

To capture all possible interconnections among the 10 traits, we included 90 items. However, to reduce survey fatigue, we implemented a planned missing data design in which participants answered all 9 statements about the outcome of happiness and a randomly-selected 5 out of 9 statements for each of the 9 other outcomes (for a total of 54 items). We used multiple imputation to handle missing data, and report pooled results based on analyses of complete datasets (see Supplemental Material, Section 1.2 for details).

We operationalized people's *beliefs about the happiness consequences* of various personality improvements by taking their responses to each of the 9 statements about how improving each of the 9 non-happiness traits would affect their happiness. In subsequent studies, we consider additional well-being outcomes beyond happiness. We operationalized *improvement centrality* as the extent to which a person believes that an improvement on a given trait (e.g., morality) would improve the other 9 traits (e.g., all traits except for morality). That is, we computed the mean of a person's beliefs about the effects of improving a given trait on all other traits.

Data Analyses

We conducted most of the analyses for all studies using R Version 3.6.3 (R Core Team, 2020). We implemented mixed-effects models using the *lme4* package (Bates et al., 2015). All mixed-effects models included random intercepts for each participant. Models predicting change goals also included random slopes for each participant and within-person centered predictor variables (all studies), whereas models that involved categorical comparisons between moral and nonmoral traits also included random intercepts for the specific input and outcome traits (Study 1 only). We used Mplus Version 8.3 (Muthén & Muthén, 1998) to fit the multiple-predictor mixed-effects model and to compute Wald tests to compare whether the three standardized within-

person regression coefficients were different from one another (Study 1 only). The multiple-predictor analysis in Study 1 excluded the outcome of happiness (because beliefs about effects on happiness was included as a predictor).

Results

Why Are People More Interested in Improving Certain Traits?

First, we examined reasons why people prioritize improving certain aspects of themselves. We tested one established explanation—the Deficit-Reduction Hypothesis—as well as additional Personality-Enabling, Identity-Continuity, and Personal Fulfillment Hypotheses.

Preregistered single-predictor mixed-effects models showed that participants were more interested in improving traits for which they perceived greater deficiencies in themselves (β = 0.25, 95% CI [0.20, 0.30], $p < .001$), which they believed would have more wide-ranging consequences for improving their other traits in general (β = 0.24, 95% CI [0.20, 0.28], $p < .001$), and which they believed would improve their happiness to a greater extent (β = 0.36, 95% CI [0.32, 0.41], $p < .001$).

When all three predictors were included in the model, perceived deficiencies (β = 0.18, 95% CI [0.13, 0.23], $p = .001$) and beliefs about happiness consequences (β = 0.28, 95% CI [0.23, 0.33], $p = .001$) continued to predict greater change goals; however, improvement centrality no longer had a detectable independent effect (β = 0.05, 95% CI [–0.05, 0.14], p = .325). All three regression coefficients significantly differed from one other (ps < .019). Thus, these results provide independent support for the Deficit-Reduction and Personal Fulfillment Hypotheses, but not the Personality-Enabling or Identity-Continuity Hypotheses.

Why Do People Deprioritize Moral Improvements?

Conceptually replicating Sun and Goodwin (2020), a non-preregistered analysis showed that participants were less interested in making moral improvements than nonmoral improvements, $\beta = -0.74$, 95% CI [–1.13, –0.35], $p = .006$. A non-preregistered analysis showed that people descriptively reported greater deficiencies on nonmoral traits than moral traits (i.e., the direction predicted by the Deficit-Reduction Hypothesis), but this was not statistically significant, $\beta = -0.63$, 95% CI [–1.37, 0.10], $p = .131$. This is likely a limitation of our categorical operationalization of moral relevance in this study; a limitation we address in Study 2. Importantly, however, controlling for perceived deficiencies ($\beta = 0.22$, 95% CI [0.17, 0.26], $p < .001$), people were still less interested in making moral improvements than nonmoral improvements ($\beta = -0.60$, 95% CI [–1.01, –0.19], $p = .020$). This suggests that there is room for additional explanations. We test several such explanations below.

The Personality-Enabling Hypothesis predicts that people are more interested in improving nonmoral traits because they believe that nonmoral improvements would facilitate greater personality improvements in general. In contrast, the Identity-Continuity Hypothesis predicts that people are less interested in improving moral traits because these traits have the greatest impact on their other traits, thus causing the most disruption to their personal identity. However, the analyses predicting change goals (described above) showed that perceived widespread personality improvements did not independently emerge as a general explanation for why people want to change some traits more than other traits. Moreover, participants believed that the overall effects of moral improvements on nonmoral improvements would be similar to the effects of nonmoral improvements on moral improvements, $\beta = -0.09$, 95% CI [–0.66, 0.49], $p = .781$ (controlling for current levels on the input and outcome traits: $\beta = -0.05$, 95% CI [–0.62, 0.52], $p = .865$). Therefore, it does not seem to be the case that people are less interested in

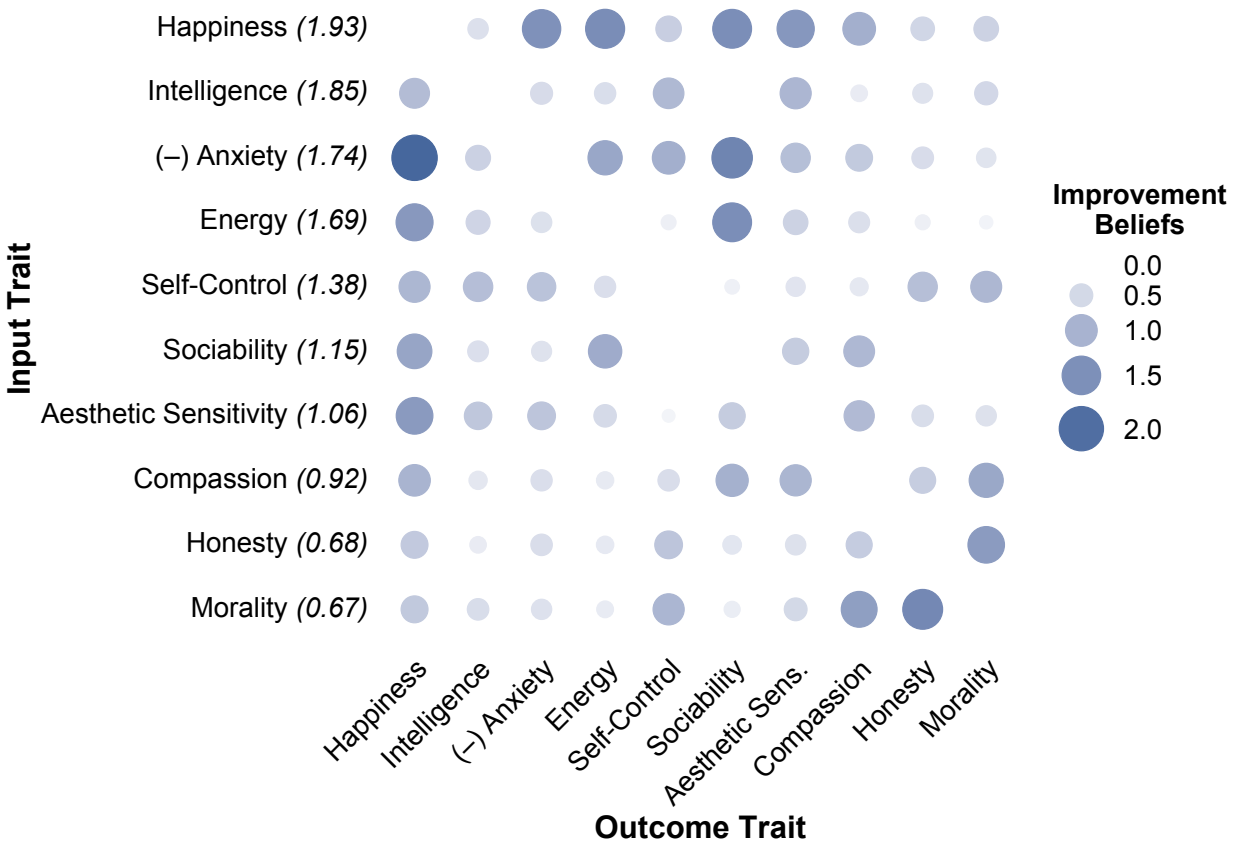
moral improvements because such improvements are expected to cause fewer positive impacts on one's personality (the Personality-Enabling Hypothesis) or, alternatively, dramatically change who they are (the Identity-Continuity Hypothesis).

The Personal Fulfillment Hypothesis predicts that people believe that moral improvements come at an absolute or relative cost to their well-being. However, participants generally believed that moral improvements would at least slightly improve their happiness (see Figure 1). This speaks against the *absolute* cost explanation; the idea that people deprioritize moral improvements because they believe that such improvements would make them *less* happy.

In contrast, we found suggestive evidence for the *relative* cost explanation. A preregistered linear mixed-effects model suggested that participants believed that nonmoral improvements would improve happiness to a greater extent than would moral improvements, although this difference was not statistically detectable, $\beta = -0.49$, 95% CI [–0.99, 0.01], $p = .098$ (controlling for current levels on input and outcome traits: $\beta = -0.51$, 95% CI [–1.03, +0.00], $p = .093$). However, within the "moral" and "nonmoral" categories, some traits were judged to have a greater impact on morality (i.e., were more morally relevant) than other traits. Thus, we also ran a non-preregistered, exploratory analysis in which we predicted beliefs about consequences for happiness from beliefs about consequences for morality, using 8 input traits (excluding happiness and morality). In other words, instead of categorically comparing moral traits to nonmoral traits, we instead conducted an analysis that modeled moral relevance as a continuous variable. This analysis suggested that participants believed that traits that were more morally relevant would improve happiness to a lesser extent, $\beta = -0.08$, 95% CI [–0.13, –0.03], $p = .014$ (see Figure S1). Studies 2a and 2b provide two preregistered, confirmatory tests of this finding.

Figure 1

Mean Beliefs about the Consequences of Each Trait Improvement on All Other Traits in Study 1



Note. The size of each circle represents the average (mean) extent to which people believe that improving each input trait (*x*-axis) would improve each outcome trait (*y*-axis). All visualized means were detectably different from 0 using a false discovery rate (FDR) corrected $p < .05$ threshold. Beliefs about the effects of sociability on honesty, self-control, and morality, and of the effect of intelligence on sociability, were not detectably different from 0. None of the input traits were judged to *decrease* levels of an outcome trait. Traits are ordered from the traits that participants most wanted to improve (Happiness) to the ones that they least wanted to improve (Morality). Mean change goals for each input trait are shown in parentheses (where –3 = much less, 0 = no change desired, 3 = much more). See Table S2 for means and standard deviations.

Discussion

In sum, participants were more interested in improving the traits that they perceived being more deficient on and that they believed would lead to greater happiness benefits. In contrast, the effect of improvement centrality was not uniquely detectable after accounting for perceived happiness consequences and perceived deficiencies on the input trait. Moreover,

although past research might suggest that moral traits are more central than nonmoral traits (Chen et al., 2016), here, we found that people believe that moral and nonmoral improvements would have similar impacts on one another. This means that neither the Personality-Enabling nor the Identity-Continuity Hypotheses can explain why people deprioritize moral improvements. Accordingly, in Studies 2a and 2b, we drop our investigation of these hypotheses to focus on further refining the Deficit-Reduction and Personal Fulfillment Hypotheses.

Interestingly, people are quite interested in becoming more intelligent and self-controlled, despite not believing they are particularly deficient on these traits nor believing that improving these traits would make them much happier. This suggests that there are likely additional well-being considerations (e.g., meaning in life, goal attainment) that may further explain peoples' trait change preferences. Thus, in Studies 2a and 2b, we assess people's beliefs about how various trait improvements would impact a broader range of valued outcomes beyond happiness.

Study 1 also helps clarify the nature of the tradeoff that people perceive between happiness and moral improvements. People do not believe that moral improvements would make them *less* happy—instead, participants generally reported that becoming more compassionate, honest, and moral would make them *happier* than they currently are. However, we find suggestive evidence that people believe that less morally relevant improvements would lead to even greater increases in happiness. Studies 2a and 2b provide a confirmatory test of this idea and examines whether it generalizes to other aspects of well-being.

One limitation of Study 1 was that it employed a categorical distinction between three "moral" and seven "nonmoral" traits. However, traits vary in the degree to which they can be considered morally relevant. The measurement imprecision involved in this categorical operationalization of moral vs. nonmoral traits might explain why people did not report being

significantly less deficient on moral traits than nonmoral traits. Moreover, there are many other moral traits beyond compassion, honesty, and morality. To allow for a more continuous measure of moral relevance and to sample a more comprehensive range of traits, in Studies 2a and 2b, we examine 22 input traits.

Studies 2a and 2b

In Studies 2a and 2b, we assess people's beliefs about how changing 22 input traits would affect morality and five aspects of well-being: day-to-day happiness, meaning in life, social connectedness, social status, and goal attainment. The expanded set of outcomes allows us to address two primary questions. First, which aspects of well-being most strongly predict people's personality change goals? Second, does the finding of a relative—but not absolute—tradeoff between morality and happiness generalize to other aspects of well-being?

We made two substantive changes between Studies 2a and 2b (additional changes are noted in the Method section). First, the definition of morality presented to participants in Studies 1–2a mentioned willingness to make self-sacrifices (see Appendix B). It is possible that the mention of self-sacrifice (a) caused participants to be less interested in improving their morality and (b) caused participants to believe that there is a tradeoff between morality and well-being. We therefore modified our definition of morality in Study 2b to eliminate any mention of self-sacrifice. Second, regardless of how high or low they are on a given trait, people might only perceive that they are "deficient" on a trait if they have lower levels than they consider to be optimal. Accordingly, in Study 2b, we examine a second indicator of perceived deficiencies: how far people feel they are from the optimal level on each trait. Because Studies 2a and 2b produced extremely similar results, we report their results together below. For concision, we report figures

for Study 2b (which we consider to be the more definitive study) in the main text, but parallel figures for Study 2a are reported in the supplement.

Note that before conducting Study 2b, we conducted a third, less comprehensive online study, with a smaller sample and more limited set of outcome variables (Supplemental Study, n = 110). The Supplemental Study ruled out the Malleability Hypothesis and generally continued to show that people perceived a relative tradeoff between morality and well-being (even when self-sacrifice was removed from the definition of morality). However, there were some findings that differed from Study 2a; for example, participants in the Supplemental Study were not detectably less interested in improving moral traits (p = .076). As a result, we decided that it would be best to conduct a full in-person replication of Study 2a, with a larger sample size and all outcome variables (i.e., Study 2b). We report the Supplemental Study results in full in Section 5 of the Supplemental Material.

Method

Participants

We recruited participants for Studies 2a and 2b from a Behavioral Lab at a business school in the U.K. The participant pool was drawn from university students, staff, and individuals from the general populace. We aimed to recruit 300 participants per study. We believed that this sample size would suffice because it would provide double the number of responses for each belief item as did Study 1 (because Study 2 involved a complete data design, as opposed to a planned missing data design).

Study 2a. Participants typically visit the Behavioral Lab in person. However, due to the COVID-19 pandemic, all studies were conducted remotely when this study was conducted. Participants were paid £5 for completing an online survey that took approximately 25–30

minutes to complete (median completion time = 27.2 minutes). 342 surveys were submitted. As preregistered, we excluded 38 participants who failed either of the two comprehension checks. The Behavioral Lab staff also flagged a potential duplicate account, which we excluded (this was not preregistered). The final sample comprised 303 participants (124 men, 172 women, 1 preferred to self-describe, 6 preferred not to say) who ranged in age from 18 to 70 years (M = 31.95, SD = 12.17). Participants reported being based in the U.K. (n = 229), U.S. (n = 12), or another country (n = 62) and identified as Asian/Asian British (n = 134), White/White British (n = 104), Black/Black British (n = 18), Other or Multiple (n = 42), or did not report their ethnicity (n = 5).

Study 2b. Participants visited the Behavioral lab in person. Participants were paid £10 for completing an online survey that took approximately 25–40 minutes to complete (actual completion times were not properly logged in this in-lab study, as the first page of the Qualtrics survey was often opened ahead of the participants' arrival). Participants who participated in Study 2a were not allowed to sign up for Study 2b. 321 participants started the study. As preregistered, we excluded 15 participants who failed either of the two comprehension checks more than once and 5 participants who did not submit the survey (n = 5). The final sample comprised 301 participants (110 men, 187 women, 2 preferred to self-describe, 2 preferred not to say) who ranged in age from 18 to 79 years (M = 32.78, SD = 12.65). Participants identified as Asian/Asian British (n = 121), White/White British (n = 102), Black/Black British (n = 28), Other or Multiple (n = 47), or did not report their ethnicity (n = 3).

Procedure

In both Studies 2a and 2b, participants viewed the definitions of 6 outcome traits (happiness, meaning, morality, goal attainment, social status, and social connectedness; see

Appendix B for definitions), before completing a comprehension check in which they were required to correctly match up each outcome with its definition. They then viewed the definitions of 22 input traits (see Appendix B), before completing a comprehension check in which they were required to correctly match up a subset of 7 or 8 traits with their definitions. The definition of morality was revised in Study 2b to remove a possible confound (described above), but all other definitions were the same across the two studies. In Study 2a, those who failed either comprehension check were automatically screened out of the survey. In Study 2b, those who failed a comprehension check were allowed to restart the study once, but if they failed again, they were paid and allowed to leave.

Participants who passed both comprehension checks proceeded to the rest of the survey. In Study 2a, participants reported, in a counterbalanced order, (a) their beliefs about how improving each of the 22 input traits would change each of the 6 outcomes (i.e., "improvement beliefs") and (b) their current levels on and change goals for the input traits and outcomes. In Study 2b, participants completed the measures in the following fixed order: (a) change goals for the input traits and outcomes, (b) improvement beliefs, and (c) current and optimal levels for the input traits and outcomes. In other words, unlike Study 1 (which assessed the perceived interconnections between all 10 traits), in Studies 2a and 2b, we focused on understanding beliefs about how improvements on 22 input traits would change 6 outcomes (which were distinct from the input traits). See Section 1.1 of the Supplemental Material for detailed justifications of the 22 input traits and 6 outcomes.

Measures

Change Goals. Input traits included 8 traits from Study 1 (excluding happiness and morality) and 14 new traits (see Appendix B and Figure 4). This expanded set of input traits

increases statistical power and enables more comprehensive coverage of the trait space. Participants reported their change goals using 1 item per input trait, using the same measure as in Study 1. For consistency of interpretation, the negative traits (anxiety, depression, and exploitativeness) were reverse-scored, such that higher scores indicate goals to become *less* anxious, depressed, or exploitative. Thus, higher scores on all change goals indicate a greater desire to *improve* each respective trait. We also assessed change goals for the outcome traits, which were used for supplemental analyses and are described in the Supplemental Material (see Section 4.2).

Perceived Deficiencies. In Study 2a, we operationalized perceived deficiencies in terms of low current levels. In Study 2b, we used two operationalizations of perceived deficiencies: Low current levels (same as Study 2a) and suboptimal levels.

Low Current Levels (Studies 2a and 2b). Because we assessed many input traits, for brevity, we assessed current levels on each trait with 1 item ("To what extent are you someone who is [trait]?"). Participants responded on an 11-point scale anchored by 0 = *Not at all*, 5 = *Moderately*, and 10 = *Completely*. We reversed the resulting scores for all traits except for anxiety, depression, and exploitativeness, such that higher levels reflect greater perceived deficiencies (i.e., less desirable levels) of each trait.

Suboptimal Levels (Study 2b only). In Study 2b, we assessed participants' optimal level on each trait using one item per trait (e.g., "What level of honesty would be optimal (best, well-suited, and adaptive) for you?"), using the same anchors as for the measures of current levels. We computed a "suboptimality" index by subtracting current levels from optimal levels (after reversing both current and optimal levels for anxiety, depression, and exploitativeness). Thus, positive scores correspond to a person believing that they are currently below their optimal level

of a trait, whereas negative scores correspond to a person believing that they are currently above their optimal level of a trait.

Beliefs about Consequences of Personality Improvements. The six outcomes were happiness, meaning in life, social connectedness, social status, goal attainment, and morality. There are many different conceptualizations of what constitutes a good life (DeYoung & Tiberius, 2022; Marsh et al., 2020; Oishi & Westgate, 2022; Ryff, 1989; Seligman, 2012). Because there is no consensual model of well-being, we used our expert judgment to select outcomes that seemed to be most central or prototypically valued, while minimizing the redundancy among different outcomes (for details, see Section 1.1 of the Supplemental Material).

As in Study 1, participants imagined a change that would make a noticeable difference in their daily life. They then reported their beliefs about how improving each of the 22 input traits (e.g., "If I became more [efficient], I would:") would change each of the six outcomes. Responses were made on a 7-point scale (–3 = much less, 0 = same amount, +3 = much more) with appropriately-worded anchors for each outcome (e.g., –3 = *Become much less socially connected than I currently am*, 0 = *Remain as socially connected as I currently am*, and +3 = *Become much more socially connected than I currently am*). For the input traits of anxiety, depression, and exploitativeness, the items were phrased in terms of becoming *less* anxious, depressed, and exploitative (respectively). In other words, positive scores correspond to a person believing that improving on a trait would lead to improved outcomes, whereas negative scores correspond to a person believing that improving on a trait would lead to worse outcomes. Participants answered all 132 items (22 input traits × 6 outcomes).

Beliefs about consequences for morality were used to assess moral relevance. That is, the more people believe that improving the trait would improve their own morality, the more "morally relevant" it is. Beliefs about consequences for goal attainment are distinct from change goals in that the former captures participants' beliefs about how improving a trait would help them to achieve their various goals, rather than participants' goals to change a given trait.

Results

Revisiting the Deficit-Reduction Hypothesis

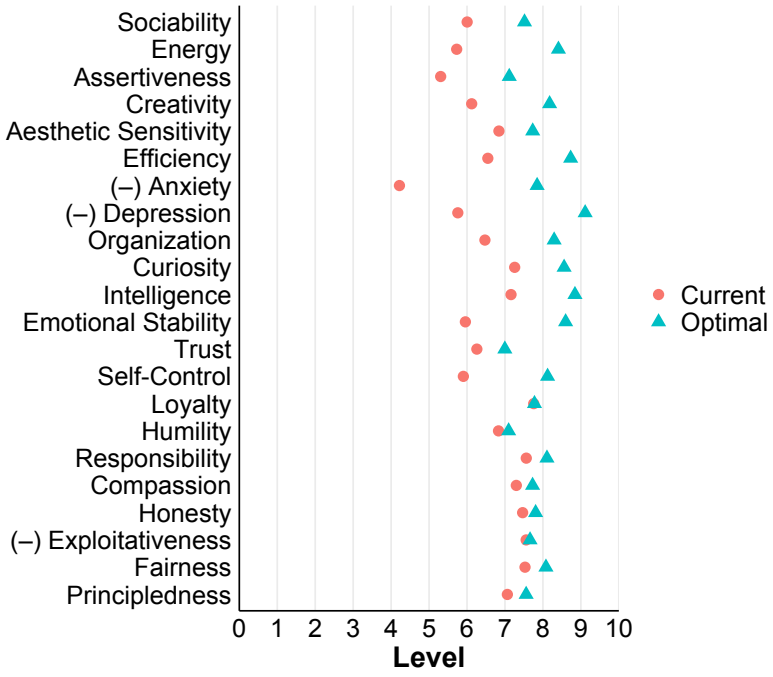
Study 1 showed that people are more interested in improving traits that they perceive being more deficient on. However, the Deficit-Reduction Hypothesis also predicts that people perceive fewer deficiencies on moral (vs. nonmoral) traits, and we only found directional support for this idea in Study 1. We provide a stronger test of this using a continuous measure of moral relevance in Studies 2a and 2b. Moreover, people may only perceive a "deficit" to the extent that their current level is below their optimal level. We test this alternative operationalization of perceived deficiencies in Study 2b.

In both Studies 2a and 2b, we find that people are less interested in making more morally relevant improvements (see Table 2, Single-Predictor Models). Non-preregistered analyses also show that people perceive fewer deficiencies for more morally relevant traits, whether perceived deficiencies is operationalized as low current levels (Study 2a: $\beta = -0.11$, 95% CI [–0.14, –0.08], $p < .001$; Study 2b: $\beta = -0.13$, 95% CI [–0.16, –0.10], $p < .001$) or suboptimal levels (Study 2b: $\beta = -0.12$, 95% CI [–0.15, –0.09], $p < .001$; see Figure 2). Importantly, however, in a non-preregistered analysis controlling for low levels in Study 2a and for both low levels and suboptimal levels in Study 2b, people are still less interested in improving on more morally relevant traits (see Table 2 and Figure 3). This analysis also shows that people are also more

interested in improving on traits that they perceive greater deficiencies on—particularly when operationalized as suboptimal levels (rather than low levels per se, as low levels was no longer a significant predictor when suboptimal levels was included in the same model in Study 2b).

Figure 2

Discrepancy Between Trait and Optimal Levels in Study 2b



Note. Traits are ordered from least morally relevant (top) to most morally relevant (bottom).

In sum, these analyses support the Deficit-Reduction Hypothesis. Going beyond Study 1 and past work (Sun & Goodwin, 2020), Study 2b also suggests that perceptions of suboptimality are especially important. Nonetheless, controlling for perceived deficiencies, people are still less interested in making moral improvements. This suggests that additional explanations are at play. Having already ruled out the Personality-Enabling and Identity-Continuity Hypotheses in Study 1, as well as the Malleability Hypothesis in the Supplemental Study (see Section 5), we proceed to clarify the Personal Fulfillment Hypothesis.

Table 2

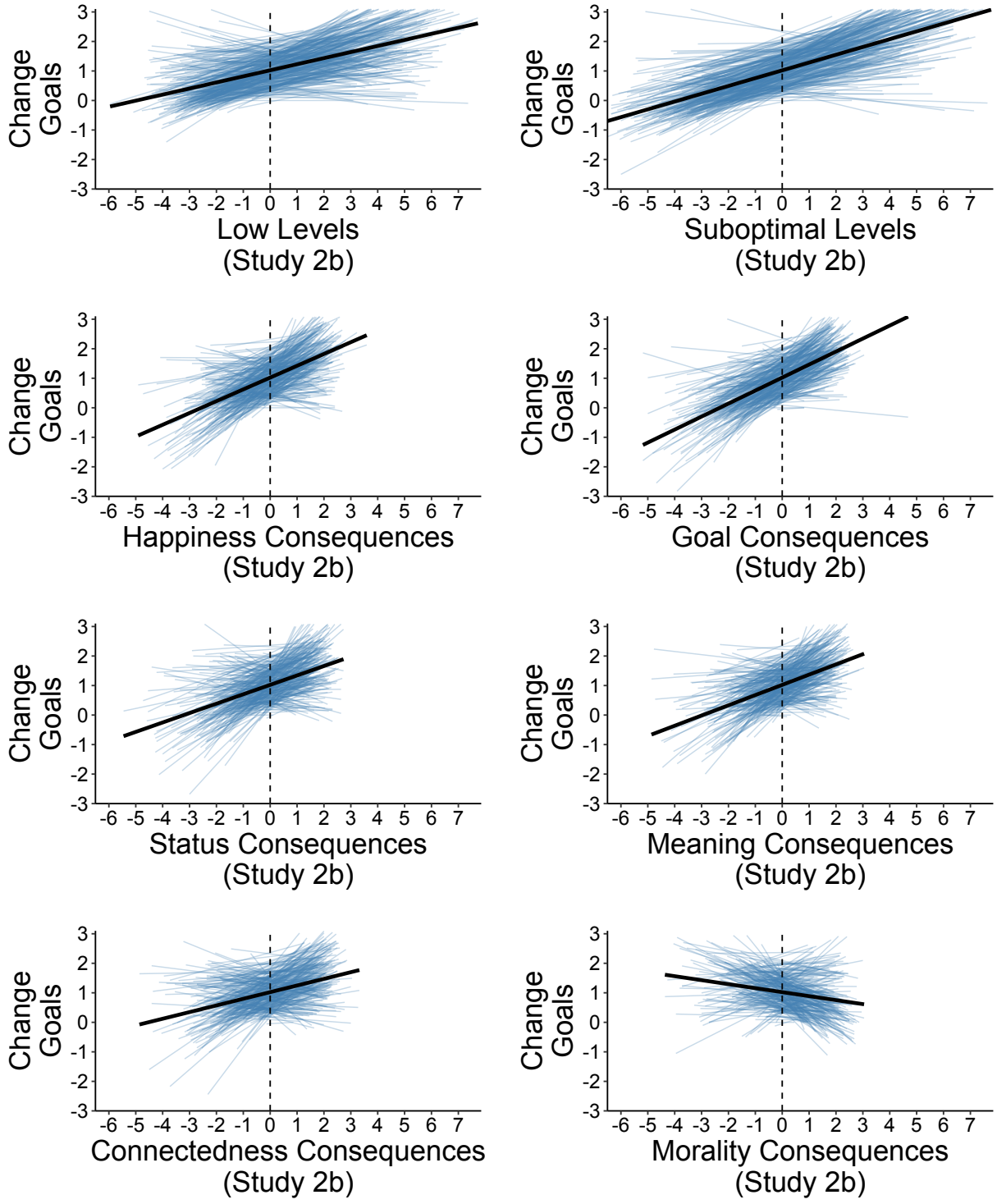
Predicting Change Goals from Perceived Deficiencies and Beliefs About Consequences for Morality and Well-Being Outcomes (Studies 2a and 2b)

	Single-Predictor Models				Multiple-Predictor Model		
Predictor	β	95% CI	*p*	τ^2	β	95% CI	*p*
Deficit-Reduction Hypothesis							
Study 2a							
Low levels					0.27	[0.23, 0.31]	< .001
Morality					–0.07	[–0.10, –0.04]	< .001
Study 2b							
Low levels					–0.00	[–0.05, 0.04]	.865
Suboptimal levels					0.50	[0.46, 0.55]	< .001
Morality					–0.06	[–0.09, –0.04]	< .001
Personal Fulfillment Hypothesis							
Study 2a							
Goal attainment	0.46	[0.42, 0.50]	< .001	.044	0.30	[0.27, 0.33]	< .001
Happiness	0.39	[0.36, 0.43]	< .001	.055	0.18	[0.14, 0.21]	< .001
Meaning	0.29	[0.25, 0.33]	< .001	.071	0.09	[0.06, 0.12]	< .001
Low levels	0.28	[0.24, 0.32]	< .001	.021	0.17	[0.14, 0.20]	< .001
Social status	0.27	[0.23, 0.30]	< .001	.075	0.12	[0.09, 0.14]	< .001
Connectedness	0.19	[0.15, 0.23]	< .001	.046	–0.01	[–0.03, 0.02]	.561
Morality	–0.11	[–0.14, –0.08]	< .001	.043	–0.02	[–0.05, 0.00]	.078
Study 2b							
Suboptimal levels	0.51	[0.48, 0.54]	< .001	.007	0.23	[0.19, 0.28]	< .001
Goal attainment	0.48	[0.45, 0.51]	< .001	.03	0.25	[0.23, 0.28]	< .001
Happiness	0.40	[0.37, 0.44]	< .001	.054	0.10	[0.07, 0.13]	< .001
Low levels	0.37	[0.33, 0.40]	< .001	.013	0.10	[0.06, 0.13]	< .001
Meaning	0.32	[0.29, 0.36]	< .001	.047	0.09	[0.06, 0.11]	< .001
Social status	0.30	[0.27, 0.34]	< .001	.053	0.12	[0.09, 0.14]	< .001
Connectedness	0.23	[0.20, 0.27]	< .001	.042	–0.02	[–0.04, 0.01]	.164
Morality	–0.13	[–0.17, –0.10]	< .001	.042	–0.04	[–0.06, –0.02]	< .001

Note. τ^2 = slope variance. For Study 2a, we fit the seven-predictor model in three steps: Perceived deficiencies (i.e., low levels; Step 1; R^2 = .15), adding beliefs about happiness (Step 2; R^2 = .28), and adding beliefs about consequences for the other five outcomes (Step 3; R^2 = .42). For Study 2b, we fit the eight-predictor model in three steps: Perceived deficiencies (i.e., low levels and suboptimal levels; Step 1; R^2 = .33), adding beliefs about happiness (Step 2; R^2 = .38), and adding beliefs about consequences for the other five outcomes (Step 3; R^2 = .47). Predictors are ordered from the most positive to most negative effects in the multiple-predictor model. For each study, predictors are ordered from the most positive to most negative effects in the single-predictor model.

Figure 3

Within-Person Associations Between Beliefs About Change Consequences and Change Goals (Study 2b)



Note. In these spaghetti plots, each colored line represents one person's slope, and the black line shows the average within-person effect. All predictor variables (*x*-axis) are within-person centered. See Table 1 for the estimates from single- and multiple-predictor multilevel models. See Figure S1 for parallel figure for Studies 1 and 2a.

Clarifying the Personal Fulfillment Hypothesis

Study 1 showed that people are more interested in improving traits that they believe would increase their happiness to a greater extent. Similarly, Gander and Wagner (2024) showed that people are more interested in improving traits that they believe are better for their "well-being" (but did not examine how people weigh different aspects of well-being). Here, we examine the relative importance of beliefs about consequences for happiness versus other aspects of well-being, as well as morality.

Relative Importance of Various Well-Being Outcomes. First, we fit a set of mixed-effects models in which we predicted change goals from perceived deficiencies and beliefs about consequences for morality and each of the five well-being outcomes (one predictor per model). As expected, in both Studies 2a and 2b, participants were more interested in improving the traits for which they perceived greater deficiencies (i.e., lower levels or suboptimal levels) in themselves (see Table 2). They were also more interested in improving traits that they believed would improve their goal attainment, happiness, meaning in life, social status, and social connectedness to a greater extent, but were less interested in improving more morally relevant traits (see Table 2 and Figure 3). Thus, these findings continue to support the Personal Fulfillment Hypothesis and show that it extends beyond the hedonic outcome of happiness.

Second, we fit a seven- (Study 2a) or eight- (Study 2b) predictor mixed effects model. In these multiple-predictor models, out of the belief predictors, beliefs about goal attainment most strongly predicted change goals in both studies (see Table 2). In both studies, beliefs about consequences for happiness, meaning, and social status also predicted greater change goals. In both studies, there was no detectable unique effect of beliefs about consequences for connectedness. Finally, beliefs about morality consequences uniquely predicted lower change

goals in Study 2b (but did not have a detectable effect in Study 2a). Thus, these results remain broadly supportive of the Personal Fulfillment Hypothesis, but suggest that beliefs about consequences for goal attainment are particularly important.

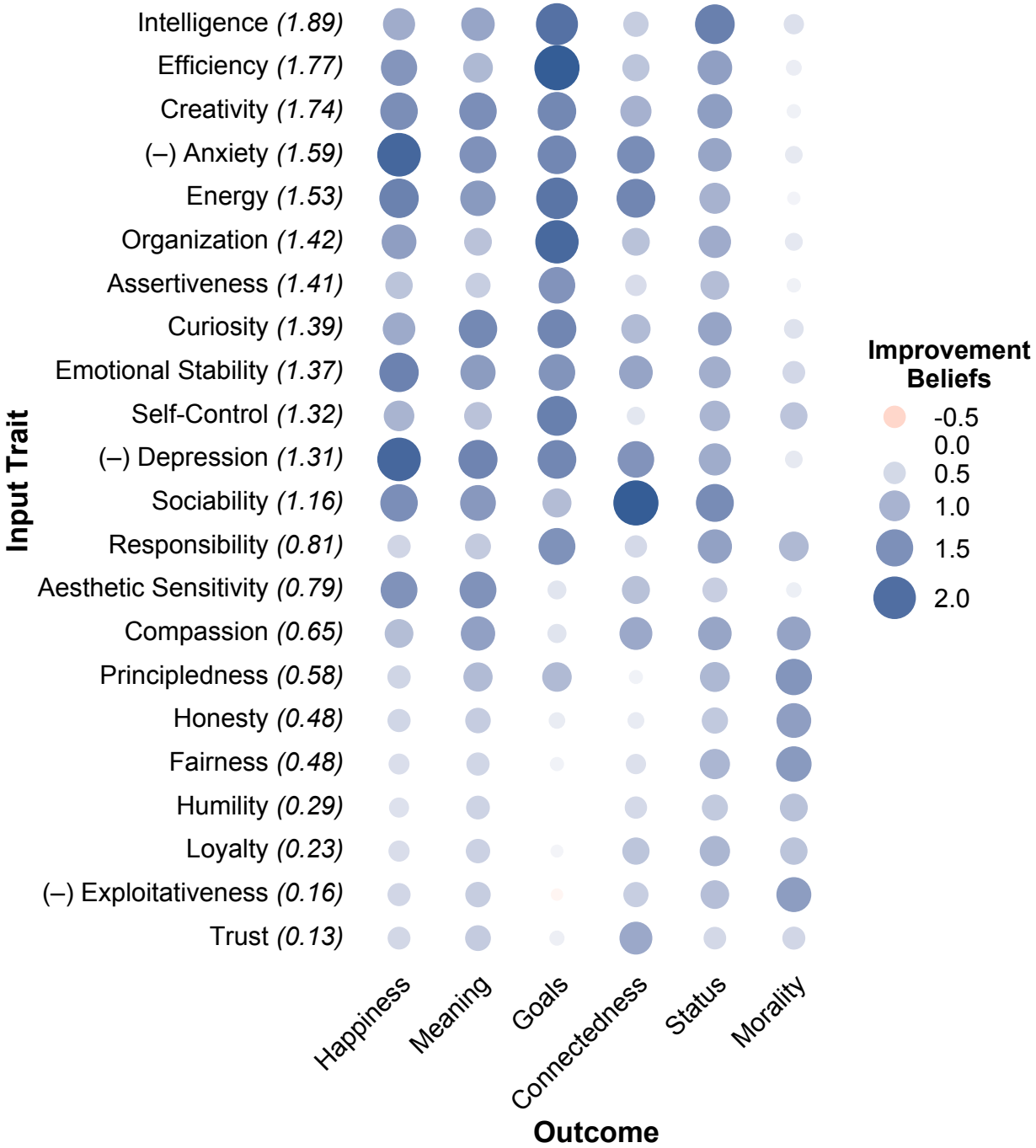
Do People Perceive Tradeoffs Between Morality and Well-Being?

Finally, Study 1 suggested that people believe that more morally relevant trait improvements would increase happiness to a lesser extent than nonmoral improvements. Is this also true for other aspects of well-being? To test this, we ran linear mixed-effects models predicting beliefs about the consequences of various trait improvements for each of the five well-being dimensions from beliefs about the consequences for morality (i.e., moral relevance).

As shown in Figure 4 (see also Figure S3, Table S5, and Table S6), participants believed that almost all trait improvements would increase each well-being dimension to some extent. Interestingly, supplemental analyses show that people even expect well-being benefits of improving traits they report being optimal or super-optimal on (see Figure S4). This suggests that people rarely perceive absolute well-being costs to improving their morality.

Figure 4

Mean Beliefs about the Consequences of Each Trait Improvement on All Outcomes in Study 2b



Note. The size of each circle represents the average (mean) extent to which people believe that improving each input trait (*x*-axis) would improve each outcome (*y*-axis). Input traits are ordered from the traits that participants most wanted to improve (Intelligence) to the ones that they least wanted to improve (Trust). All means were detectably different from 0 at an FDR-corrected $p < .05$ level, except for the effects of humility on goal attainment and sociability on morality (not visualized). Mean change goals for each input trait are shown in parentheses (where the response scale corresponded to –3 = much less, –2 = less, –1 = slightly less, 0 = no change desired, 1 = slightly more, 2 = more, 3 = much more). See Table S6 for means and standard deviations. See Figure S3 for parallel figure for Study 2a.

Table 3

Predicting Beliefs About Consequences for Well-Being Outcomes from Beliefs About Consequences for Morality (Studies 2a and 2b)

	Study 2a			Study 2b		
Outcome	β	95% CI	*p*	β	95% CI	*p*
Goal attainment	**–0.18**	**[–0.21, –0.14]**	**< .001**	**–0.17**	**[–0.21, –0.13]**	**< .001**
Happiness	**–0.08**	**[–0.12, –0.04]**	**< .001**	**–0.12**	**[–0.15, –0.08]**	**< .001**
Social connectedness	–0.01	[–0.05, 0.03]	.726	**–0.05**	**[–0.09, –0.01]**	**.008**
Meaning	–0.00	[–0.04, 0.04]	.91	0.03	[–0.01, 0.07]	.192
Social status	**0.07**	**[0.03, 0.11]**	**< .001**	**0.05**	**[0.02, 0.09]**	**.005**

Note. Effects that were statistically detectable at a $p < .05$ threshold are shown in **boldface**. Predictors are ordered from the most negative to the most positive effects. See Figure S2 for spaghetti plots. Supplemental analyses showed that these results generally held after controlling for the smaller deficiencies that people perceive for more morally relevant traits (see Table S7).

However, participants believed that less morally relevant improvements would increase goal attainment and happiness to a greater extent than would more morally relevant improvements (see Table 3 and Figure S2). Participants in Study 2b also believed that less (vs. more) morally relevant improvements would increase social connectedness to a greater extent, but this effect was much smaller compared to the effects for goal attainment and happiness, and did not emerge in Study 2a. We did not find detectable effects of traits' moral relevance on beliefs about consequences for meaning in life. Interestingly, social status showed the opposite pattern: Participants intuited that *more* morally relevant improvements would increase social status to a greater extent. Thus, people have nuanced beliefs about how moral improvements would affect various aspects of their well-being. These findings suggest that people are less interested in becoming more moral because they especially perceive tradeoffs for their goal attainment and happiness (rather than the other well-being outcomes we examined).

Discussion

In sum, Studies 2a and 2b support both the Deficit-Reduction Hypothesis and the Personal Fulfillment Hypothesis. In line with the Deficit-Reduction Hypothesis, people tended to

perceive smaller deficiencies on moral traits, and they were more interested in improving traits that they perceived greater deficiencies on—especially when deficiencies were operationalized as having suboptimal levels (vs. low levels per se; Study 2b). However, even after accounting for perceived deficiencies, people were still less interested in improving moral traits. In line with the Personal Fulfillment Hypothesis, people were more interested in improving traits that they believed would improve various aspects of their well-being (independently of perceived deficiencies). Specifically, people are especially interested in improving the traits that they believe would help them achieve their goals. They are quite interested in improving the traits that will increase their day-to-day happiness, and slightly interested in improving the traits that will increase their social status and meaning in life. Thus, Studies 2a and 2b illustrate the value of considering a wider range of valued outcomes when considering why people want to change some personality traits more than others.

Studies 2a and 2b also continue to clarify the nature of the well-being–morality tradeoff that people perceive. People do not perceive *absolute* costs of moral improvements on any of the aspects of well-being that we assessed. Instead, people think that improving morally-relevant traits will generally improve their happiness, goal attainment, meaning in life, social connectedness and social status. However, people hold nuanced beliefs about the comparative advantage of improving more vs. less morally relevant traits for different outcome variables. Specifically, people expect that moral (vs. nonmoral) improvements would lead to relatively smaller increases in happiness and goal attainments, similar increases in meaning in life and social connectedness, and smaller increases in social status.

General Discussion

Past research emphasizes the importance of morality to personal identity and impression formation (Brambilla et al., 2021; Chen et al., 2016; Goodwin et al., 2014; Hartley et al., 2016; Strohminger & Nichols, 2014). However, people downplay morality when it comes to personal improvement (Sun & Goodwin, 2020). We find that this is not only because people already see themselves as being highly (or near-optimally) moral, but also because people mainly seek to improve the traits that they believe will increase their ability to accomplish their goals and, to a lesser extent, their day-to-day happiness. Although people think that improving their moral traits would increase their well-being, they believe that nonmoral improvements would increase their well-being to an even greater extent. Thus, we conclude that people deprioritize moral improvements because they prioritize their well-being more highly than being moral. In the context of self-improvement, people seem to primarily be asking, "What's in it for me?"

Personal Change is for Personal Fulfillment

Our results are broadly in line with recent work showing that people are more interested in improving traits that they believe contribute more positively to their well-being (Gander & Wagner, 2024). However, illustrating the value of assessing specific aspects of well-being, Studies 2a–2b showed that beliefs about consequences for goal attainment were more important than beliefs about happiness consequences in explaining which traits people are more interested in changing. This helps to explain why people seek to improve competence-related traits such as intelligence, efficiency, and self-control, even though they do not believe that such improvements would substantially increase their day-to-day happiness.

Supplemental analyses (see Supplemental Material, Section 4.2) additionally showed that beliefs about how trait improvements would impact various outcomes (e.g., happiness, meaning) more positively predicted change goals to the extent that people wanted to improve these

respective outcomes. For example, someone who is more interested in increasing their social status tends to be more interested in improving the traits that they believe would lead to larger increases in their social status (compared to someone who is less interested in increasing their social status). These findings suggest that any theory of motivations for personality change should account for the fact that there are many values beyond happiness, and that people place different weights on various values.

Clarifying the Perceived Connections Between Well-Being and Morality

Our findings also clarify the nature of the tradeoff that people perceive between well-being and morality. We found that people believe that moral improvements would slightly *improve* all well-being dimensions that we measured. Thus, people do not believe that becoming more moral would come at an *absolute cost* to their well-being—even when they see themselves as already being at or above the optimal levels of moral traits (see Figure S4). Instead, they believe that moral improvements would have differentiated relative well-being impacts: More morally relevant improvements would lead to relatively greater increases in status, relatively smaller gains in happiness and goal attainment, and similar gains in meaning in life and social connectedness. However, because personality change goals were most strongly predicted by people's beliefs about the consequences for goal attainment and happiness—the two outcomes that people believe would be least efficiently improved by morality—people deprioritize more morally relevant improvements. In other words, perceptions of *relative* opportunity costs can help explain why people are less interested in becoming more moral.

These findings also speak to the ongoing debate about whether happiness and meaning are differentially associated with morality (Baumeister et al., 2013; Choi et al., 2017; Dwyer et al., 2017; Hofmann et al., 2014). Past theory and research purports that meaningful lives are

characterized by moral principles, societal contribution, giving, and self-transcendence, whereas happy lives are characterized by comfort, being a taker, personal satisfaction, and self-enhancement (Baumeister et al., 2013; Dwyer et al., 2017; Huang & Yang, 2022; Oishi & Westgate, 2022; Prinzing et al., 2022). This might imply that people would believe that moral improvements would (a) increase meaning to a greater extent than would nonmoral improvements and (b) increase meaning to a greater extent than happiness. Our results, however, show that people believe that moral and nonmoral improvements would increase their meaning in life to a similar extent. This might be because people have many nonmoral sources of meaning in life, including vocations, relationships, intellectual and aesthetic pursuits, and daily routines (Heintzelman & King, 2019; Wolf, 1997). Still, people did intuit that moral improvements would lead to smaller improvements in their happiness than would nonmoral improvements. In other words, although people perceive no particular advantage of moral improvements for increasing their meaning in life, they perceive a greater tradeoff between morality and happiness than between morality and meaning.

Limitations and Future Directions

Our studies find that people don't want to be more moral in part because they believe that nonmoral improvements would be more beneficial for their well-being. Future research could use experiments to assess the accuracy of these beliefs. For example, do nonmoral improvements actually increase happiness to a greater extent than do moral improvements? If so, could such knowledge be used to encourage greater moral improvement? Alternatively, even if moral improvements increase well-being, it may be that explicitly seeking well-being benefits from moral improvements feels distasteful (cf. Barasch et al. 2014).

Although we assessed a wide range of outcome measures, we did not examine how various trait improvements were expected to impact feelings of authenticity. Given that people see their "true self" (i.e., who they are deep down inside) as being moral (Bench et al., 2015; Strohminger et al., 2017), it is possible that moral improvements may make people feel more authentic.

Additionally, whereas people do not particularly want to improve their moral traits, they might be relatively more alarmed about the prospect of moral deterioration (Molouki & Bartels, 2017; Strohminger & Nichols, 2014). Thus, whereas the Personal Fulfillment Hypothesis helps to explain which traits people most want to *improve*, there may be additional explanations for which deteriorations people are most motivated to *avoid*.

Finally, as mentioned, supplemental analyses (see Supplemental Material, Section 4.2) suggest that the specific ways in which the Personal Fulfillment Hypothesis manifests for a particular person depends on how they weigh various dimensions of well-being. In a similar vein, different cultures likely place different weights on various dimensions of well-being. For instance, Markus and Kitayama (1991) have argued that people in interdependent cultures seek to obtain competence to socially connect and meet the expectations of significant others rather than further their own advancement. If so, then differences in the weights that different cultures place on various dimensions of well-being might help to explain cross-cultural differences in personality change goals (Baranski et al., 2021). Future research should test whether this is the case.

Conclusion

Why don't people particularly want to be more moral? Why do people want to change some traits more than other traits? Three studies support both the Deficit-Reduction Hypothesis

and the Personal Fulfillment Hypothesis: People primarily want to improve traits that they believe they are more deficient on, as well as traits that will most increase their ability to achieve their goals and their day-to-day happiness. Although people believe that moral improvements would slightly increase these well-being outcomes, they believe that nonmoral improvements would increase them to an even greater extent. Thus, beliefs about the relative well-being benefits of moral vs. nonmoral improvements may help to explain why people deprioritize moral improvements. These results point to the centrality of well-being for understanding why people want to improve themselves and clarify the perceived tradeoffs between well-being and morality.

Author Contributions

[Author 1] and [Author 2] conceptualized and designed the studies. [Author 2] collected the data. [Author 1] and [Author 2] planned the preregistered data analyses. [Author 1] analyzed and visualized the data. [Author 1] drafted the manuscript, and [Author 2] provided critical revisions.

References

Aquino, K., & Reed, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, *83*, 1423–1440. https://doi.org/10.1037/0022-3514.83.6.1423

Baranski, E., Gardiner, G., Lee, D., & Funder, D. C. (2021). Who in the world is trying to change their personality traits? Volitional personality change among college students in six continents. *Journal of Personality and Social Psychology*, *121*, 1140–1156. https://doi.org/10.1037/pspp0000389

Barasch, A., Levine, E. E., Berman, J. Z., & Small, D. A. (2014). Selfish or selfless? On the signal value of emotion in altruistic behavior. *Journal of Personality and Social Psychology*, *107*(3), 393–413. https://doi.org/10.1037/a0037207

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

Baumeister, R. F., & Exline, J. J. (1999). Virtue, personality, and social relations: Self-control as the moral muscle. *Journal of Personality*, *67*(6), 1165–1194. https://doi.org/10.1111/1467-6494.00086

Baumeister, R. F., Vohs, K. D., Aaker, J. L., & Garbinsky, E. N. (2013). Some key differences between a happy life and a meaningful life. *The Journal of Positive Psychology*, *8*(6), 505–516. https://doi.org/10.1080/17439760.2013.830764

Bench, S. W., Schlegel, R. J., Davis, W. E., & Vess, M. (2015). Thinking about change in the self and others: The role of self-discovery metaphors and the true self. *Social Cognition*, *33*(3), 169–185. https://doi.org/10.1521/soco.2015.33.3.2

765 Bigman, Y. E., & Tamir, M. (2016). The road to heaven is paved with effort: Perceived effort
766 amplifies moral judgment. *Journal of Experimental Psychology: General*, *145*(12), 1654.
767 https://doi.org/10.1037/xge0000230
768 Brambilla, M., Sacchi, S., Rusconi, P., & Goodwin, G. P. (2021). The primacy of morality in
769 impression development: Theory, research, and future directions. In B. Gawronski (Ed.),
770 *Advances in Experimental Social Psychology* (Vol. 64, pp. 187–262). Academic Press.
771 https://doi.org/10.1016/bs.aesp.2021.03.001
772 Chen, S. Y., Urminsky, O., & Bartels, D. M. (2016). Beliefs about the causal structure of the self-
773 concept determine which changes disrupt personal identity. *Psychological Science*,
774 *27*(10), 1398–1406. https://doi.org/10.1177/0956797616656800
775 Choi, J., Catapano, R., & Choi, I. (2017). Taking stock of happiness and meaning in everyday
776 life: An experience sampling approach. *Social Psychological and Personality Science*,
777 *8*(6), 641–651. https://doi.org/10.1177/1948550616678455
778 De Freitas, J., Cikara, M., Grossmann, I., & Schlegel, R. (2018). Moral goodness is the essence
779 of personal identity. *Trends in Cognitive Sciences*, *22*(9), 739–740.
780 https://doi.org/10.1016/j.tics.2018.05.006
781 DeYoung, C. G., & Tiberius, V. (2022). Value fulfillment from a cybernetic perspective: A new
782 psychological theory of well-being. *Personality and Social Psychology Review*, 3–27.
783 https://doi.org/10.1177/10888683221083777
784 DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects
785 of the Big Five. *Journal of Personality and Social Psychology*, *93*(5), 880–896.
786 https://doi.org/10.1037/0022-3514.93.5.880

787 Dwyer, R. J., Dunn, E. W., & Hershfield, H. E. (2017). Cousins or conjoined twins: How
788 different are meaning and happiness in everyday life? *Comprehensive Results in Social*
789 *Psychology*, *2*(2–3), 199–215. https://doi.org/10.1080/23743603.2017.1376580
790 Effron, D. A., & Miller, D. T. (2015). Do as I say, not as I've done: Suffering for a misdeed
791 reduces the hypocrisy of advising others against it. *Organizational Behavior and Human*
792 *Decision Processes*, *131*, 16-32. https://doi.org/10.1016/j.obhdp.2015.07.004
793 Epley, N., & Dunning, D. (2000). Feeling "holier than thou": Are self-serving assessments
794 produced by errors in self-or social prediction? *Journal of Personality and Social*
795 *Psychology*, *79*(6), 861–875. https://doi.org/10.1037/0022-3514.79.6.861
796 Erlandsson, A., Wingren, M., & Andersson, P. A. (2020). Type and amount of help as predictors
797 for impression of helpers. PloS one, 15(12), e0243808.
798 https://doi.org/10.0.5.91/journal.pone.0243808
799 Furr, R. M., Prentice, M., Hawkins Parham, A., & Jayawickreme, E. (2022). Development and
800 validation of the Moral Character Questionnaire. *Journal of Research in Personality*, *98*,
801 104228. https://doi.org/10.1016/j.jrp.2022.104228
802 Gai, P. J., & Bhattacharjee, A. (2022). Willpower as moral ability. *Journal of Experimental*
803 *Psychology: General*, *151*, 1999–2006. https://doi.org/10.1037/xge0001169
804 Gander, F., & Wagner, L. (2024). Which positive personality traits do people want to
805 change?. *European Journal of Personality*, *38*(4), 657–676.
806 https://doi.org/10.1177/08902070231211957
807 Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., &
808 Gough, H. G. (2006). The international personality item pool and the future of public-

domain personality measures. *Journal of Research in Personality*, *40*, 84-96. https://doi.org/10.1016/j.jrp.2005.08.007

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

Gray, K., & Wegner, D. M. (2009). Moral typecasting: Divergent perceptions of moral agents and moral patients. *Journal of Personality and Social Psychology, 96*(3), 505–520. https://doi.org/10.1037/a0013748

Hartley, A. G., Furr, R. M., Helzer, E. G., Jayawickreme, E., Velasquez, K. R., & Fleeson, W. (2016). Morality's centrality to liking, respecting, and understanding others. *Social Psychological and Personality Science*, *7*(7), 648–657. https://doi.org/10.1177/1948550616655359

Heintzelman, S. J., & King, L. A. (2019). Routines and meaning in life. *Personality and Social Psychology Bulletin*, *45*(5), 688–699. https://doi.org/10.1177/0146167218795133

Hofmann, W., Meindl, P., Mooijman, M., & Graham, J. (2018). Morality and self-control: How they are intertwined and where they differ. *Current Directions in Psychological Science*, *27*(4), 286–291. https://doi.org/10.1177/0963721418759317

Hofmann, W., Wisneski, D. C., Brandt, M. J., & Skitka, L. J. (2014). Morality in everyday life. *Science*, *345*(6202), 1340–1343. https://doi.org/10.1126/science.1251560

Huang, M., & Yang, F. (2022). Self-transcendence or self-enhancement: People's perceptions of meaning and happiness in relation to the self. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001297

Hudson, N. W., & Fraley, R. C. (2016). Changing for the better? Longitudinal associations between volitional personality change and psychological well-being. *Personality and Social Psychology Bulletin*, *42*(5), 603–615. https://doi.org/10.1177/0146167216637840

Hudson, N. W., & Roberts, B. W. (2014). Goals to change personality traits: Concurrent links between personality traits, daily behavior, and goals to change oneself. *Journal of Research in Personality*, *53*, 68–83. https://doi.org/10.1016/j.jrp.2014.08.008

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

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*, 224–253. https://doi.org/10.1037/0033-295X.98.2.224

Marsh, H. W., Huppert, F. A., Donald, J. N., Horwood, M. S., & Sahdra, B. K. (2020). The well-being profile (WB-Pro): Creating a theoretically based multidimensional measure of well-being to advance theory, research, policy, and practice. *Psychological Assessment*, *32*, 294–313. https://doi.org/10.1037/pas0000787

Molouki, S., & Bartels, D. M. (2017). Personal change and the continuity of the self. *Cognitive Psychology*, *93*, 1–17. https://doi.org/10.1016/j.cogpsych.2016.11.006

Monin, B., & Jordan, A. H. (2009). The dynamic moral self: A social psychological perspective. In *Personality, identity, and character: Explorations in moral psychology* (pp. 341–354).

Muthén, L. K., & Muthén, B. O. (1998). *Mplus User's Guide* (Eighth). Los Angeles, CA: Muthén & Muthén.

O'Connor, K., Effron, D. A., & Lucas, B. J. (2020). Moral cleansing as hypocrisy: When private acts of charity make you feel better than you deserve. *Journal of Personality and Social Psychology*, *119*(3), 540–550. https://doi.org/10.1037/pspa0000195

Oishi, S., & Westgate, E. C. (2022). A psychologically rich life: Beyond happiness and meaning. *Psychological Review*, *129*, 790–811. https://doi.org/10.1037/rev0000317

Olivola, C. Y., & Shafir, E. (2013). The martyrdom effect: When pain and effort increase prosocial contributions. *Journal of Behavioral Decision Making*, *26*(1), 91-105. https://doi.org/10.1002/bdm.767

Peterson, C., & Seligman, M. E. (2004). *Character strengths and virtues: A handbook and classification*. Oxford University Press.

Prinzing, M., De Freitas, J., & Fredrickson, B. L. (2022). The ordinary concept of a meaningful life: The role of subjective and objective factors in third-person attributions of meaning. *The Journal of Positive Psychology*, *17*(5), 639–654. https://doi.org/10.1080/17439760.2021.1897866

Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*, *57*, 1069–1081. https://doi.org/10.1037/0022-3514.57.6.1069

Seligman, M. E. P. (2012). *Flourish: A visionary new understanding of happiness and well-being*. Simon and Schuster.

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, *113*, 117–143. https://doi.org/10.1037/pspp0000096

Strohminger, N., Knobe, J., & Newman, G. (2017). The true self: A psychological concept distinct from the self. *Perspectives on Psychological Science*, *12*(4), 551–560. https://doi.org/10.1177/1745691616689495

Strohminger, N., & Nichols, S. (2015). Neurodegeneration and identity. *Psychological Science*, *26*(9), 1469–1479. https://doi.org/10.1177/0956797615592381

Sun, J., & Goodwin, G. P. (2020). Do people want to be more moral? *Psychological Science*, *31*(3), 243–257. https://doi.org/10.1177/0956797619893078

Sun, J., Wilt, J., Meindl, P., Watkins, H., & Goodwin, G. P. (2024). How and why people want to be more moral. *Journal of Personality*, *92*(3), 907–925. https://doi.org/10.1111/jopy.12812

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, *72*(2), 271–324.

Tappin, B. M., & McKay, R. T. (2017). The illusion of moral superiority. *Social Psychological and Personality Science*, *8*, 623–631. https://doi.org/10.1177/1948550616673878

Thielmann, I., & de Vries, R. E. (2021). Who wants to change and how? On the trait-specificity of personality change goals. *Journal of Personality and Social Psychology*, *121*, 1112–1139. https://doi.org/10.1037/pspp0000304

Wolf, S. (1997). Happiness and meaning: Two aspects of the good life. *Social Philosophy and Policy*, *14*(1), 207–225. https://doi.org/10.1017/S0265052500001734

Appendix A

Trait Definitions and Items Used to Measure Current Levels in Study 1

Trait	Definition	Items used to measure current levels
Moral traits		
Morality	A **moral** person does the right thing, leads an ethical lifestyle, and is willing to make personal sacrifices for the benefit of others and the greater good.	Does the right thing[6] Is willing to make personal sacrifices for the benefit of others and the greater good[6] Leads an ethical lifestyle[6]
Honesty	An **honest** person consistently tells the truth and refuses to cheat or steal in order to get ahead.	Cheats to get ahead[5] [r] Consistently tells the truth[4]
Compassion	A **compassionate** person feels sympathy for others and is helpful and unselfish.	Feels little sympathy for others[1] [r] Is helpful and unselfish with others[1]
Nonmoral traits		
Happiness	A **happy** person experiences a lot of positive emotion, does not experience very much negative emotion, and feels satisfied with their life as a whole.	Experiences a lot of positive emotion[6] Experiences a lot of negative emotion[6] [r] Feels satisfied with their life as a whole[6]
Anxiety	An **anxious** person feels tense and worries a lot.	Can be tense[1] Worries a lot[1]
Energy level	An **energetic** person is full of vitality and is rarely tired.	Is full of energy[1] Often feels tired[5] [r]
Sociability	A **sociable** person is outgoing, talkative, and likes to spend time with other people.	Is outgoing, sociable[1] Is talkative[1]
Self-control	A **self-controlled** person has the willpower to resist temptations and is able to work effectively toward long-term goals.	Is good at resisting temptation[4] Is able to work effectively toward long-term goals[4]
Aesthetic sensitivity	An **aesthetically sensitive** person appreciates beauty (e.g., in art, music, literature, and nature).	Values art and beauty[1] Enjoys the beauty of nature[5]
Intelligence	An **intelligent** person learns quickly and can handle a lot of information.	Can handle a lot of information[2] Learns quickly[5]

Note. We aimed to use items that aligned with the definitions that participants saw. Where possible, we selected items from previously validated measures of these traits, or items from the IPIP database. In a few cases, we used ad hoc measures. [r] denotes a reverse-scored item. Superscripts denote the source of the item: [1]Big Five Inventory - Revised (Soto & John, 2017), [2]Big Five Aspect Scales (DeYoung et al., 2007), [3]Moral Characteristics Questionnaire (Furr et al., 2022), [4]Brief Self-Control Scale (Tangney et al., 2004), [5]IPIP database (Goldberg et al., 2006), [6]Ad-hoc item (based on the definition).

Appendix B

Definitions for Traits in Studies 2a and 2b

Trait	Definition
Input Traits	
Intelligence	An **intelligent** person learns quickly and can handle a lot of information.
Energy	An **energetic** person is full of vitality and is rarely tired.
Efficiency	An **efficient** person gets things done quickly and doesn't like to waste time.
Creativity	A **creative** person comes up with original ideas and finds clever ways to do things.
Anxiety	An **anxious** person feels tense and worries a lot.
Sociability	A **sociable** person is outgoing, talkative, and likes to spend time with other people.
Emotional Stability	An **emotionally stable** person keeps their emotions under control and does not experience many mood swings.
Depression	A **depressed** person often feels sad, is easily discouraged, and tends to be insecure.
Organization	An **organized** person is systematic and likes to keep things in order.
Self-Control	A **self-controlled** person has the willpower to resist temptations and is able to work effectively toward long-term goals.
Assertiveness	An **assertive** person takes charge and is willing to express their opinions and preferences in social situations.
Curiosity	An **intellectually curious** person likes to engage with abstract ideas and is curious about many different things.
Aesthetic Sensitivity	An **aesthetically sensitive** person appreciates beauty (e.g., in art, music, literature, and nature).
Responsibility	A **responsible** person is committed to meeting their duties and can be depended on by others.
Compassion	A **compassionate** person feels sympathy for others and is helpful and unselfish.
Principledness	A **principled** person has an internally consistent set of moral rules that they strictly follow.
Honesty	An **honest** person consistently tells the truth and refuses to cheat or steal in order to get ahead.
Humility	A **humble** person is unassuming and sees themselves as an ordinary person who is not entitled to special treatment.
Fairness	A **fair** person is impartial and unprejudiced in how they treat others.
Exploitativeness	An **exploitative** person takes advantage of others for their own personal gain.
Loyalty	A **loyal** person shows firm and consistent support to a person or institution and does not shift their allegiances easily.
Trust	A **trusting** person assumes the best about other people and is not suspicious about their intentions.
Outcomes	
Day-to-day happiness	**Day-to-day happiness** describes the extent to which a person experiences a lot of positive emotion and not very much negative emotion in their daily life.
Meaning in life	**Meaning in life** describes the extent to which a person has a sense of purpose, believes that their life has value, and can make sense of the things that happen in their life.
Goal attainment	**Goal attainment** describes the extent to which a person is achieving the various goals they have for themselves.
Social connectedness	**Social connectedness** describes the extent to which a person has supportive relationships, feels close to others, and rarely feels lonely.
Social status	**Social status** describes the extent to which a person is respected and admired by other people.
Morality	**Morality** describes the extent to which a person does the right thing, leads an ethical lifestyle, and [Study 1 and Study 2a: is willing to make personal sacrifices; Study 2b and Supplemental Study: acts] for the benefit of others and the greater good.

Table of Contents

1. Additional Methodological Details

1.1. Trait Selection

In Study 1, we selected the three moral traits of morality, honesty, and compassion because we were interested in people's beliefs about moral improvements in general, as well as specific dimensions of morality. Although there are many dimensions of moral character (e.g., Furr et al., 2022; Walker & Frimer, 2007), we focused on honesty and compassion for this initial investigation because they are two of the most prototypical moral virtues and capture the well-established distinction between integrity and benevolence (Goodwin et al., 2014; Moore et al., 2019).

We chose the 7 nonmoral traits based on the following considerations. We included *happiness* because we were specifically interested in what people believe about the happiness implications of various trait improvements. To provide some representativeness in our sampling of nonmoral traits, we aimed to have coverage across four of the Big Five domains of extraversion (*sociability*, *energy level*), conscientiousness (*self-control*), neuroticism (*anxiety*), and openness to experience (*aesthetic sensitivity*, *intelligence*). Note that the agreeableness domain is represented by the moral trait of *compassion*. For each of these domains, we chose the facets (i.e., narrower traits within each trait domain) that were either the most highly-prioritized within each domain (*anxiety*, *sociability*; Sun & Goodwin, 2020) or had the potential to be seen as general "enabling" factors (*energy level*, *intelligence*, *self-control*). We also included one trait that we thought participants might see as being less interconnected with other trait improvements (*aesthetic sensitivity*). See Appendix A for the definitions that participants saw, as well as the items used to measure these traits.

In Studies 2a and 2b, we expanded the set of input traits to increase statistical power and the comprehensiveness of the coverage of the trait space. We used the list of 21 traits from Sun and Goodwin (2020) as a starting point, because this list of traits included the key facets of the Big Five traits (Soto & John, 2017), as well as a number of core moral character traits (Furr et al., 2022). We selected 18 of these traits, removing morality (because we decided to include it only as an outcome trait), purity (because some have argued that this construct is conceptually vague; Gray et al., 2022), and anger (because we decided that it was sufficiently covered by emotional stability). We also included four additional traits that were not represented in this set: self-control and intelligence (which were both included in Study 1), as well as principledness and humility.

In Studies 2a and 2b, we also aimed to consider a broader range of outcomes that people might value, and that might therefore be relevant to explaining peoples' personality change goals, beyond "happiness." After considering the features of well-being that are featured in various philosophical and psychological theories of well-being, we settled on the following six outcomes.

As in Study 1, we included happiness, but relabelled it as *day-to-day happiness* (the extent to which a person experiences a lot of positive emotion and not very much negative emotion in their daily life) to emphasize a more momentary–experiential conceptualization. For concision, we continue to refer to this as "happiness." We included *meaning in life* (the extent to which a person has a sense of purpose, believes that their life has value, and can make sense of the things that happen in their life), because it is most commonly contrasted with happiness (e.g., Baumeister et al., 2013; Dwyer et al., 2017; Oishi & Westgate, 2022).

We included *social connectedness* (the extent to which a person has supportive relationships, feels close to others, and rarely feels lonely), because the dimension of relatedness is featured in many taxonomies of well-being (e.g., Butler & Kern, 2016; Ryff, 1989), and is often posited as being a basic need (Baumeister & Leary, 1995; Deci & Ryan, 2012). We included *social status* (the extent to which a person is respected and admired by other people) because we thought that it could plausibly explain peoples' interest in improving traits that are relevant to competence and agency (e.g., intelligence and self-control). We also thought of social connectedness and status as being complementary outcomes, in that social connectedness should be more relevant to warmth-based traits, whereas social status should be more relevant to competence- and agency-based traits.

Beyond these specific outcomes, desire satisfaction perspectives on well-being suggest that a good life involves fulfilling one's most valued goals, whatever they may be (DeYoung & Tiberius, 2022). Thus, we included *goal attainment* (the extent to which a person is achieving the various goals they have for themselves) as a potentially more general explanation of personality change goals.

Finally, we included *morality* (the extent to which a person does the right thing, leads an ethical lifestyle, and is willing to make personal sacrifices for the benefit of others and the greater good) to examine whether people perceive tradeoffs between morality and other valued outcomes beyond happiness. See Appendix B for the definitions that participants saw for the input traits and outcomes in Studies 2a and 2b.

1.2. Details of Planned Missing Data Design

In Study 1, the survey contained 10 blocks of 9 items each, in which each block contained items about a given outcome trait (e.g., how changes in each of the 9 other traits would

change a person's happiness). Participants answered all 9 statements about the outcome of happiness (i.e., to what extent improving an input trait would affect their happiness) and a randomly-selected 5 out of 9 statements for each of the 9 other outcomes. The order of blocks and the order of the items within each block were randomized.

Because of this planned missingness, participants in Study 1 did not respond to all items that were relevant to computing the improvement centrality of each trait. To handle this missingness, we used multiple imputation. As preregistered, we used the R package *mice* (Buuren & Groothuis-Oudshoorn, 2011) to generate 20 imputed datasets. The imputation model was based only on the 90 belief variables (1 row per participant, 1 column per belief variable). We then computed the improvement centrality composites for these complete datasets, conducted analyses on these 20 complete datasets, and pooled the estimates from these 20 analyses using Rubin's (1987) rules. We used this pooling approach for all analyses that involved the improvement centrality variable, as well as for analyses predicting change goals and happiness consequences from beliefs about morality consequences. For the latter two analyses, we randomly selected one complete dataset to visualize the respective within-person effects in Figure S1.

In Studies 2a and 2b, participants completed all 132 items. The survey contained 6 blocks of 22 items, in which each block contained items about a given outcome trait (e.g., how changes in each of the 22 input traits would change a person's meaning in life). After responding to 22 statements about a given outcome, participants moved onto the next outcome (e.g., goal attainment). The order of blocks and the order of the items within each block were randomized.

2. Descriptive Statistics for Studies 1–2b

Table S1

Descriptive Statistics (Means and SDs) for Change Goals and Improvement Centrality (Study 1)

Traits	Change Goals	Perceived Deficiencies
Happiness	1.93 (1.03)	2.5 (0.94)
Intelligence	1.85 (1.07)	1.91 (0.72)
(–) Anxiety	1.74 (1.52)	3.14 (1.18)
Energy Level	1.69 (1.08)	2.97 (1)
Self-Control	1.38 (1.17)	2.12 (0.71)
Sociability	1.15 (1.17)	3.25 (1.11)
Aesthetic Sensitivity	1.06 (1.1)	1.89 (0.83)
Compassion	0.92 (1.05)	2.15 (0.83)
Honesty	0.68 (1.02)	1.71 (0.62)
Morality	0.67 (0.97)	1.96 (0.54)

Table S2

Means (SDs) for Beliefs About Consequences of Various Trait Improvements (Study 1)

	Outcome Trait										
Input Trait	Happiness	Intelligence	Anxiety	Energy	Self-Control	Sociability	Aesthetic Sens.	Compassion	Honesty	Morality	M_{row}
Happiness		0.40 (0.97)	1.48 (1.57)	1.54 (1.02)	0.64 (1.17)	1.52 (1.03)	1.39 (1.13)	1.06 (1.16)	0.54 (0.99)	0.59 (0.98)	1.02
Intelligence	0.88 (1.32)		0.46 (1.30)	0.43 (0.96)	0.93 (1.18)	0.06 (1.21)	0.96 (1.13)	0.25 (1.04)	0.37 (0.82)	0.51 (1.07)	0.54
Anxiety	2.10 (1.06)	0.60 (0.93)		1.17 (1.25)	1.06 (1.36)	1.66 (1.21)	0.85 (1.16)	0.70 (0.98)	0.45 (0.96)	0.35 (0.80)	0.99
Energy	1.38 (1.09)	0.56 (0.98)	0.39 (1.42)		0.20 (1.15)	1.52 (1.12)	0.58 (1.04)	0.41 (0.88)	0.20 (0.74)	0.16 (0.60)	0.6
Self-Control	0.96 (1.21)	0.85 (1.06)	0.79 (1.37)	0.43 (1.07)		0.19 (1.05)	0.35 (0.86)	0.30 (0.97)	0.84 (1.09)	0.95 (1.11)	0.63
Sociability	1.20 (1.26)	0.41 (1.16)	0.38 (1.67)	1.11 (1.13)	-0.17 (1.25)		0.67 (0.98)	0.94 (1.10)	0.10 (0.91)	0.13 (0.94)	0.53
Aesthetic Sens.	1.35 (1.11)	0.74 (1.05)	0.76 (1.27)	0.49 (0.90)	0.14 (0.82)	0.67 (0.91)		0.90 (1.12)	0.44 (0.91)	0.39 (0.85)	0.65
Compassion	0.99 (1.29)	0.31 (0.93)	0.43 (1.28)	0.28 (0.90)	0.44 (1.06)	1.03 (1.00)	0.97 (1.09)		0.66 (1.06)	1.16 (1.07)	0.7
Honesty	0.71 (1.17)	0.25 (0.73)	0.44 (1.35)	0.28 (0.77)	0.76 (1.05)	0.33 (1.03)	0.39 (0.93)	0.66 (1.13)		1.34 (1.19)	0.57
Morality	0.71 (1.17)	0.44 (0.94)	0.39 (1.29)	0.26 (0.77)	0.96 (1.16)	0.24 (0.99)	0.50 (1.16)	1.29 (1.13)	1.61 (1.16)		0.71
M_{column}	1.14	0.51	0.61	0.66	0.55	0.80	0.74	0.73	0.58	0.62	

Table S3

Descriptive Statistics and Correlation Matrix for Perceived Deficiencies and Beliefs About Consequences of Personality Improvements (Study 2a)

				Intercorrelations							
Variable	*M*	SD_{WP}	SD_{BP}	1.	2.	3.	4.	5.	6.	7.	8.
1. Change goals	1.22	0.62	1.16		.29	.63	.62	.60	.62	.58	.58
2. Low levels	3.44	1.08	2.16	.28		–.01	.08	–.03	.11	.12	.08
3. Happiness beliefs	1.08	0.72	1.11	.39	.22		.78	.79	.74	.82	.82
4. Meaning beliefs	1.04	0.69	1.04	.29	.14	.42		.80	.76	.78	.79
5. Goal beliefs	1.12	0.62	1.26	.44	.2	.33	.23		.81	.79	.81
6. Connectedness beliefs	0.93	0.67	1.13	.18	.2	.35	.27	.12		.76	.76
7. Social status beliefs	1.15	0.75	1	.25	.06	.21	.19	.25	.22		.80
8. Morality beliefs	0.76	0.64	1.01	–.10	–.10	–.06	.02	–.18	.01	.07	

Note. Correlations below the diagonal are within-person; correlations above the diagonal are between-person.

Table S4

Descriptive Statistics and Correlation Matrix for Perceived Deficiencies and Beliefs About Consequences of Personality Improvements (Study 2b)

				Intercorrelations								
Variable	*M*	SD_{WP}	SD_{BP}	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. Change goals	1.02	0.48	1.16		.2	.38	.63	.55	.66	.48	.57	.52
2. Low levels	3.43	0.97	2.1	.37		.58	.01	–.05	–.02	.01	.08	.07
3. Suboptimal levels	1.42	1.02	2.29	.51	.72		.23	.15	.22	.21	.28	.16
4. Happiness beliefs	1.07	0.61	1.15	.39	.22	.42		.8	.85	.76	.73	.67
5. Meaning beliefs	1.03	0.59	1.05	.31	.14	.3	.47		.84	.75	.71	.59
6. Goal beliefs	1.09	0.56	1.26	.48	.21	.4	.4	.27		.83	.8	.67
7. Connectedness beliefs	0.88	0.58	1.13	.22	.2	.28	.39	.32	.19		.73	.64
8. Social status beliefs	1.05	0.6	1.06	.29	.09	.24	.26	.23	.29	.3		.62
9. Morality beliefs	0.62	0.49	1	–.12	–.12	–.11	–.10	.03	–.15	–.04	.06	

Note. Correlations below the diagonal are within-person; correlations above the diagonal are between-person.

Table S5

Means (SDs) for Change Goals and Predictors of Change Goals (Study 2a)

			Beliefs About Consequences For…					
Input Trait	Change goals	Low level	Happy	Meaning	Goals	Connect	Status	Morality
Intelligence	1.88	2.70	0.88	1.09	1.93	0.55	1.61	0.46
	(1.16)	(1.81)	(1.36)	(1.25)	(1.10)	(1.32)	(1.20)	(1.06)
Energy	1.86	4.52	1.73	1.25	1.99	1.71	1.18	0.33
	(1.13)	(2.23)	(1.13)	(1.25)	(1.03)	(1.23)	(1.18)	(0.91)
Efficiency	1.81	3.40	1.29	1.00	2.21	0.69	1.37	0.29
	(1.15)	(2.16)	(1.25)	(1.20)	(0.99)	(1.18)	(1.15)	(0.83)
Creativity	1.71	4.04	1.36	1.40	1.59	0.81	1.23	0.28
	(1.12)	(2.33)	(1.11)	(1.10)	(1.14)	(1.10)	(1.14)	(0.86)
(–) Anxiety	1.65	5.88	1.98	1.44	1.76	1.59	1.16	0.33
	(1.59)	(2.56)	(1.10)	(1.22)	(1.17)	(1.19)	(1.15)	(0.89)
Sociability	1.56	4.52	1.56	1.46	1.06	2.33	1.65	0.32
	(1.14)	(2.35)	(1.15)	(1.09)	(1.43)	(1.02)	(1.08)	(1.06)
Emotional Stability	1.56	4.16	1.77	1.30	1.45	1.25	1.23	0.74
	(1.25)	(2.36)	(1.15)	(1.23)	(1.21)	(1.19)	(1.16)	(1.09)
(–) Depression	1.53	4.11	2.05	1.58	1.51	1.55	1.10	0.50
	(1.53)	(2.78)	(1.14)	(1.34)	(1.25)	(1.21)	(1.16)	(1.02)
Organization	1.52	3.39	1.19	0.90	1.96	0.69	1.09	0.44
	(1.23)	(2.42)	(1.17)	(1.16)	(1.05)	(1.12)	(1.13)	(0.94)
Self-Control	1.50	3.79	1.11	0.91	1.86	0.58	1.17	1.02
	(1.26)	(2.22)	(1.28)	(1.25)	(1.13)	(1.22)	(1.14)	(1.16)
Assertiveness	1.47	4.71	0.80	0.65	1.67	0.50	1.07	0.18
	(1.21)	(2.34)	(1.25)	(1.25)	(1.15)	(1.42)	(1.44)	(1.15)
Curiosity	1.39	2.47	1.01	1.39	1.59	0.70	1.28	0.49
	(1.17)	(1.94)	(1.22)	(1.25)	(1.19)	(1.20)	(1.14)	(1.06)
Aesthetic Sens.	1.16	2.89	1.40	1.44	0.34	0.70	0.68	0.32
	(1.23)	(2.11)	(1.22)	(1.21)	(1.05)	(1.12)	(1.12)	(1.04)
Responsibility	1.07	2.40	0.58	0.82	1.41	0.55	1.32	1.08
	(1.21)	(1.89)	(1.30)	(1.10)	(1.24)	(1.10)	(1.13)	(1.08)
Compassion	0.93	2.75	0.97	1.22	0.35	1.20	1.32	1.28
	(1.18)	(2.00)	(1.35)	(1.22)	(1.18)	(1.17)	(1.20)	(1.23)
Principledness	0.87	2.99	0.59	0.91	0.83	0.30	1.17	1.56
	(1.11)	(1.90)	(1.25)	(1.12)	(1.32)	(1.14)	(1.24)	(1.24)
Honesty	0.64	2.59	0.58	0.72	0.28	0.42	0.83	1.41
	(1.10)	(1.83)	(1.22)	(1.11)	(1.15)	(1.35)	(1.40)	(1.23)
Humility	0.64	3.31	0.64	0.75	0.21	0.70	0.99	1.05
	(1.24)	(2.17)	(1.24)	(1.20)	(1.20)	(1.23)	(1.34)	(1.14)
Fairness	0.61	2.69	0.52	0.64	0.24	0.66	1.14	1.49
	(0.96)	(1.84)	(1.09)	(1.08)	(1.05)	(1.12)	(1.28)	(1.20)
(–) Exploitativeness	0.50	2.38	0.63	0.62	–0.04	0.78	0.92	1.45
	(1.26)	(2.19)	(1.20)	(1.16)	(1.22)	(1.20)	(1.35)	(1.24)
Loyalty	0.49	2.16	0.51	0.64	0.18	0.93	1.08	0.96
	(0.97)	(1.81)	(1.07)	(1.15)	(1.03)	(1.21)	(1.21)	(1.19)
Trust	0.41	3.85	0.65	0.77	0.30	1.21	0.81	0.78
	(1.14)	(2.40)	(1.32)	(1.14)	(1.25)	(1.20)	(1.30)	(1.19)

Note. All means for the belief variables were significantly different from 0 at a FDR-corrected $p < .05$ level, except for the effect of exploitativeness on goal attainment.

Table S6

Means (SDs) for Change Goals and Predictors of Change Goals (Study 2b)

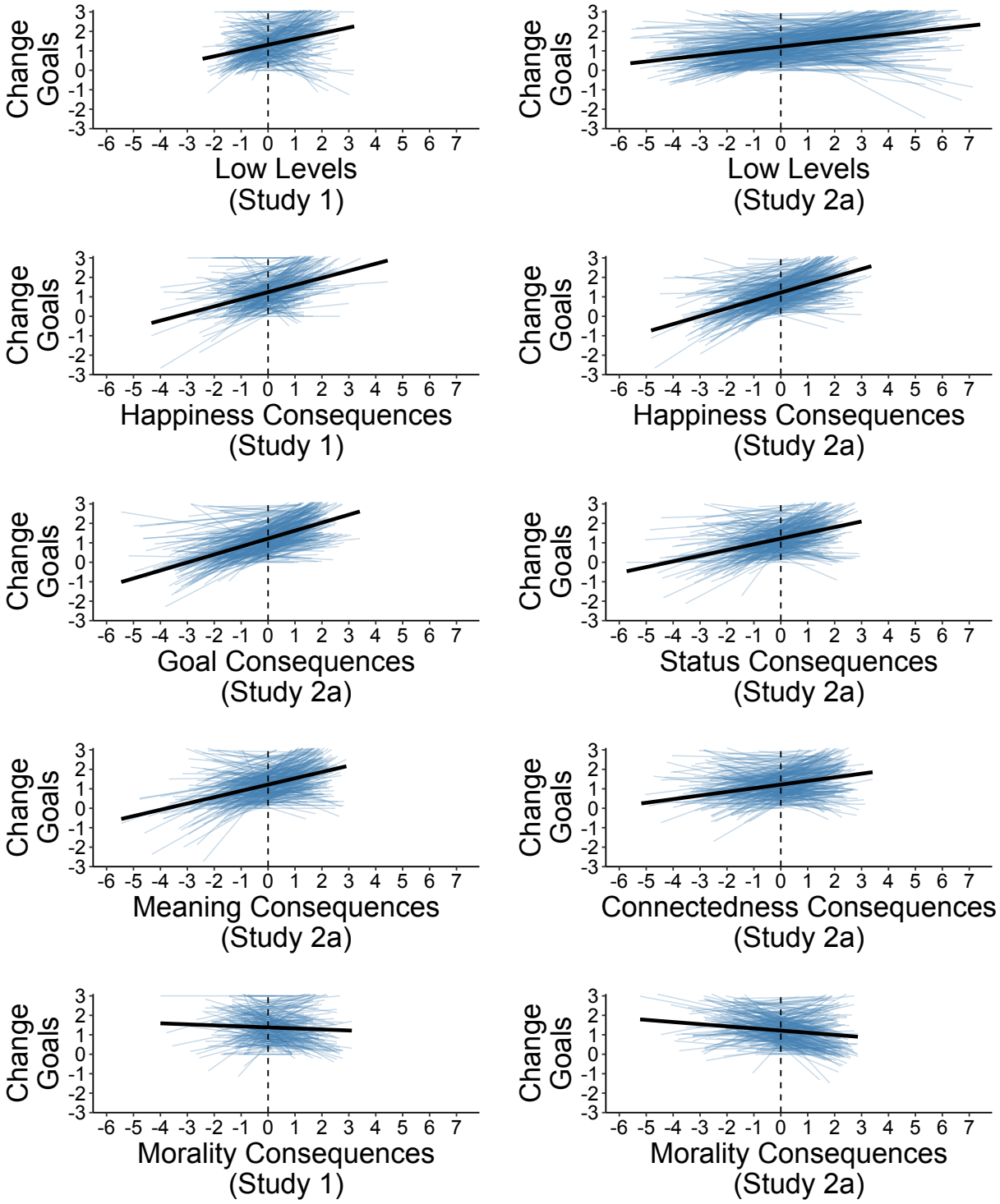
				Beliefs About Consequences For…					
Input Trait	Change goals	Low level	Subopt. level	Happy	Meaning	Goals	Connect	Status	Morality
Intelligence	1.89	2.84	1.68	1.10	1.20	1.95	0.65	1.72	0.40
	(1.08)	(1.49)	(1.85)	(1.26)	(1.27)	(1.11)	(1.27)	(1.15)	(1.08)
Efficiency	1.77	3.45	2.18	1.42	0.93	2.27	0.77	1.29	0.23
	(1.08)	(1.98)	(2.27)	(1.14)	(1.13)	(0.94)	(1.13)	(1.09)	(0.77)
Creativity	1.74	3.88	2.05	1.53	1.52	1.62	1.03	1.30	0.18
	(1.05)	(2.12)	(1.97)	(1.06)	(1.06)	(1.15)	(1.10)	(1.10)	(0.69)
Anxiety	1.59	5.78	3.63	2.11	1.48	1.63	1.54	1.20	0.29
	(1.39)	(2.85)	(3.31)	(1.12)	(1.19)	(1.24)	(1.23)	(1.14)	(0.84)
Energy	1.53	4.27	2.68	1.70	1.36	1.90	1.64	1.02	0.15
	(1.11)	(2.08)	(2.29)	(1.09)	(1.12)	(1.07)	(1.11)	(1.13)	(0.69)
Organization	1.42	3.53	1.82	1.30	0.80	2.08	0.80	1.11	0.30
	(1.10)	(2.19)	(2.42)	(1.17)	(1.09)	(1.04)	(1.13)	(1.02)	(0.74)
Assertiveness	1.41	4.69	1.81	0.78	0.64	1.45	0.45	0.87	0.18
	(1.12)	(2.22)	(2.53)	(1.23)	(1.14)	(1.18)	(1.36)	(1.37)	(1.06)
Curiosity	1.39	2.74	1.30	1.13	1.60	1.64	0.90	1.21	0.37
	(1.14)	(1.85)	(1.84)	(1.17)	(1.10)	(1.15)	(1.20)	(1.15)	(0.95)
Emotional Stability	1.37	4.04	2.64	1.70	1.33	1.44	1.21	1.08	0.51
	(1.18)	(2.22)	(2.63)	(1.19)	(1.14)	(1.11)	(1.19)	(1.08)	(0.86)
Self-Control	1.32	4.10	2.22	0.99	0.80	1.73	0.32	0.98	0.77
	(1.26)	(2.19)	(2.64)	(1.25)	(1.21)	(1.18)	(1.18)	(1.10)	(0.98)
Depression	1.31	4.24	3.36	2.10	1.66	1.63	1.45	1.10	0.29
	(1.13)	(2.87)	(3.11)	(1.15)	(1.26)	(1.23)	(1.27)	(1.16)	(0.90)
Sociability	1.16	4.00	1.51	1.52	1.38	0.88	2.26	1.56	0.06
	(1.13)	(2.22)	(2.09)	(1.17)	(1.07)	(1.42)	(1.03)	(1.18)	(1.00)
Responsibility	0.81	2.44	0.54	0.54	0.69	1.48	0.50	1.26	0.93
	(1.12)	(1.70)	(1.94)	(1.25)	(1.13)	(1.20)	(1.03)	(1.06)	(1.07)
Aesthetic	0.79	3.16	0.89	1.47	1.47	0.34	0.82	0.63	0.21
Sensitivity	(1.13)	(2.26)	(1.92)	(1.18)	(1.18)	(1.04)	(1.12)	(1.11)	(0.84)
Compassion	0.65	2.70	0.43	0.86	1.27	0.35	1.15	1.20	1.23
	(1.11)	(1.88)	(1.86)	(1.24)	(1.14)	(1.01)	(1.05)	(1.09)	(1.13)
Principledness	0.58	2.94	0.50	0.54	0.89	0.91	0.17	0.94	1.43
	(1.02)	(1.87)	(1.84)	(1.13)	(1.14)	(1.29)	(1.04)	(1.22)	(1.22)
Honesty	0.48	2.53	0.34	0.53	0.66	0.25	0.26	0.71	1.30
	(0.96)	(1.76)	(1.92)	(1.14)	(1.02)	(1.05)	(1.20)	(1.32)	(1.23)
Fairness	0.48	2.47	0.55	0.43	0.54	0.17	0.40	0.96	1.36
	(0.83)	(1.64)	(1.73)	(0.89)	(0.92)	(0.87)	(0.93)	(1.06)	(1.14)
Humility	0.29	3.17	0.26	0.39	0.57	0.11	0.50	0.70	0.81
	(1.08)	(2.04)	(2.29)	(1.19)	(1.15)	(1.07)	(1.10)	(1.35)	(1.05)
Loyalty	0.23	2.24	0.02	0.44	0.60	0.14	0.77	0.97	0.78
	(0.84)	(1.68)	(1.68)	(0.98)	(0.91)	(0.95)	(1.08)	(1.12)	(1.15)
Exploitativeness	0.16	2.44	0.10	0.53	0.66	–0.13	0.64	0.86	1.31
	(0.97)	(2.25)	(2.06)	(1.14)	(1.19)	(1.08)	(1.15)	(1.28)	(1.26)
Trust	0.13	3.74	0.74	0.52	0.68	0.20	1.15	0.51	0.53
	(1.27)	(2.12)	(2.02)	(1.28)	(1.03)	(1.09)	(1.22)	(1.30)	(1.10)

Note. All means for the belief variables were significantly different from 0 at a FDR-corrected $p < .05$ level, except for the effect of humility on goal attainment and sociability on morality.

3. Supplemental Figures for Studies 1–2b

Figure S1

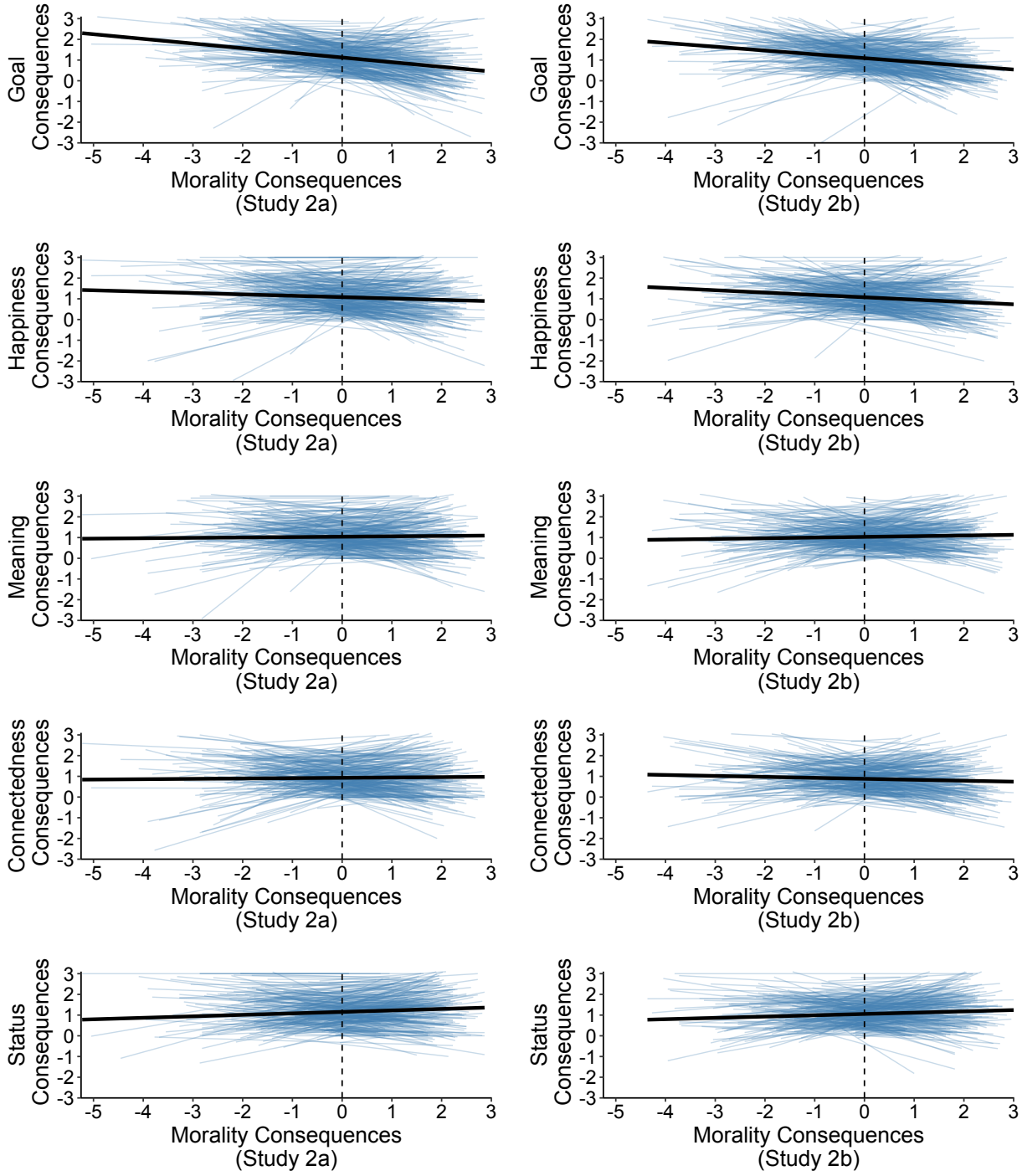
Within-Person Associations Between Beliefs About Change Consequences and Change Goals (Studies 1 and 2a)



Note. In these spaghetti plots, each colored line represents one person's slope, and the black line shows the average within-person effect. All predictor variables (*x*-axis) are within-person centered. See Table 2 for the estimates from single- and multiple-predictor multilevel models.

Figure S2

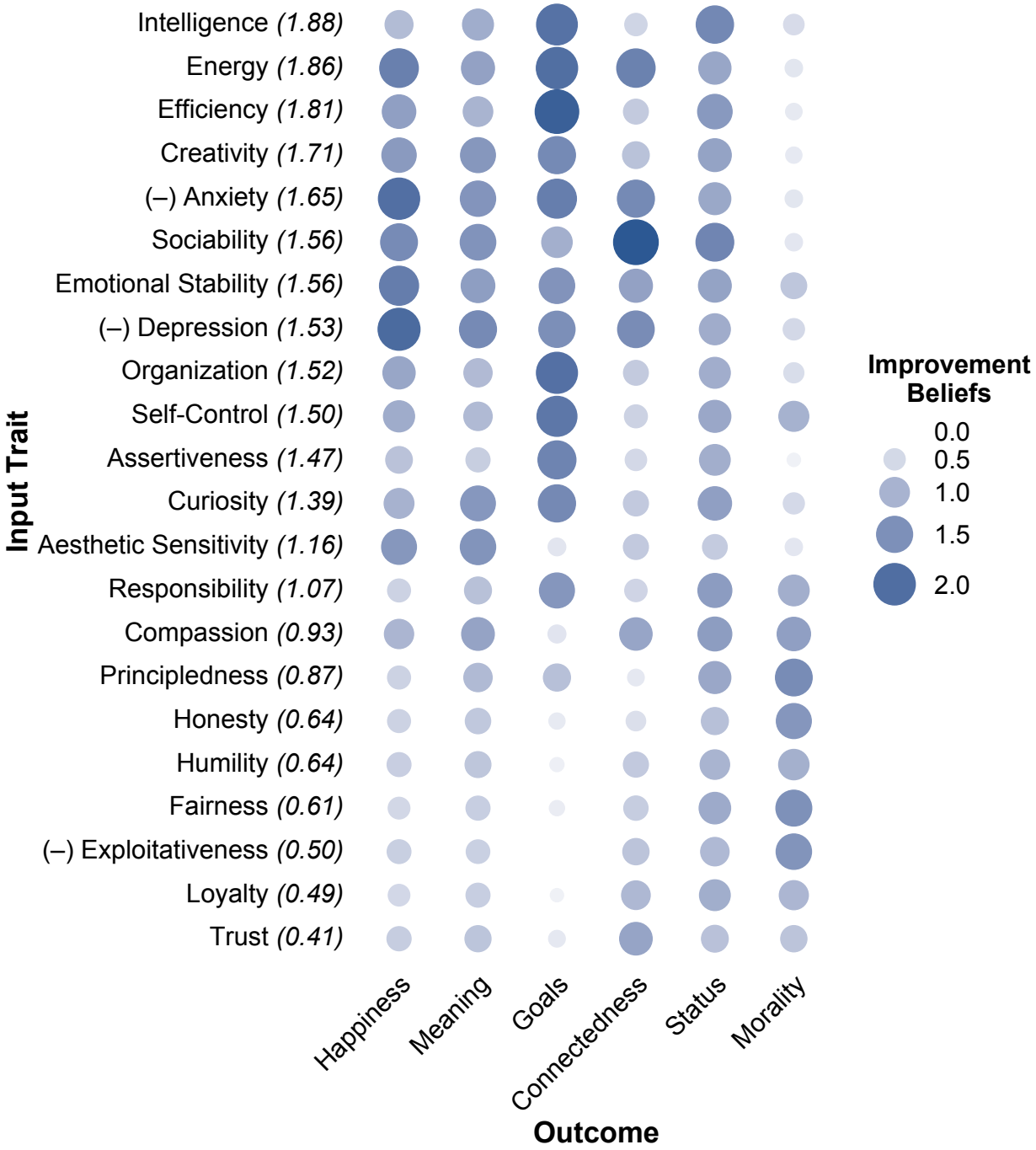
Within-Person Associations Between Beliefs About Consequences for Morality and Beliefs About Consequences for Personal Fulfillment (Studies 2a and 2b)



Note. In these spaghetti plots, each colored line represents one person's slope, and the black line shows the average within-person effect. The beliefs about morality consequences variable (*x*-axis) was within-person centered.

Figure S3

Mean Beliefs about the Consequences of Each Trait Improvement on All Outcomes in Study 2a



Note. The size of each circle represents the average (mean) extent to which people believe that improving each input trait (*x*-axis) would improve each outcome (*y*-axis). Input traits are ordered from the traits that participants most wanted to improve (Intelligence) to the ones that they least wanted to improve (Trust). All means were detectably different from 0 at an FDR-corrected $p < .05$ level, except for the effect of exploitativeness on goal attainment (not visualized). Mean change goals for each input trait are shown in parentheses (where the response scale corresponded to –3 = much less, –2 = less, –1 = slightly less, 0 = no change desired, 1 = slightly more, 2 = more, 3 = much more). See Table S5 for means and standard deviations.

Figure S4

Mean Beliefs about the Consequences of Each Trait Improvement on All Outcomes for Participants Who Were At or Above the Optimal Level of Each Trait in Study 2b

Note. The size of each circle represents the average (mean) extent to which people believe that improving each input trait (x-axis) would improve each outcome (y-axis). Input traits are ordered from the traits that participants most wanted to improve (Intelligence) to the ones that they least wanted to improve (Trust). All visualized means (121 out of 132, or 91.7%) were detectably different from 0 at an FDR-corrected $p < .05$ level. Mean change goals for each input trait are shown in parentheses (where the response scale corresponded to –3 = much less, –2 = less, –1 = slightly less, 0 = no change desired, 1 = slightly more, 2 = more, 3 = much more).

4. Supplemental Analyses for Studies 2a–2b

4.1. Can Relative Tradeoffs Between Morality and Well-Being Be Explained By Differences in Perceived Trait Deficiencies?

It is plausible that people perceive diminishing marginal returns to trait improvements, such that people imagine that the same amount of improvement in a trait would yield greater happiness benefits (for example) when they are more deficient on that trait. If so, then the relative tradeoff between morality and happiness (for example) could potentially be explained by people perceiving greater deficiencies on less morally relevant traits (rather than an inherent tradeoff between morality and happiness). Indeed, non-preregistered analyses showed that participants reported higher levels (Study 2a: $\beta = -0.11$, 95% CI [–0.14, –0.08], $p < .001$; Study 2b: $\beta = -0.13$, 95% CI [–0.16, –0.10], $p < .001$) and a smaller distance from the optimal level (Study 2b: $\beta = -0.12$, 95% CI [–0.16, –0.08], $p < .001$) for more morally relevant traits.

Accordingly, we ran another five models (preregistered in Study 2b but not Study 2a) predicting beliefs about how personality improvements would change the other five outcomes from beliefs about how personality improvements would change one's morality, controlling for perceived deficiencies (low levels in Study 2a, and both low levels and suboptimal levels in Study 2b). As shown in Table S7, the effects of beliefs about morality consequences on well-being variables was very similar. The two exceptions were that in Study 2b, the negative effect on social connectedness was no longer detectable, whereas the positive effect on meaning became detectable. In other words, we find robust evidence that people believe that making less (vs. more) morally relevant improvements would result in larger increases in their goal attainment and happiness and smaller increases in their social status, even when accounting for the smaller deficiencies that people perceive for more morally relevant traits.

Table S7

Predicting Beliefs About Consequences for Well-Being Outcomes from Perceived Deficiencies and Beliefs About Consequences for Morality (Studies 2a and 2b)

	Study 2a			Study 2b		
Outcome	β	95% CI	*p*	β	95% CI	*p*
Goal attainment						
Low level	0.17	[0.14, 0.20]	< .001	–0.20	[–0.25, –0.15]	< .001
Suboptimal level				0.52	[0.47, 0.57]	< .001
Morality	**–0.15**	**[–0.19, –0.12]**	**< .001**	**–0.12**	**[–0.15, –0.08]**	**< .001**
Happiness						
Low level	0.2	[0.16, 0.23]	< .001	–0.18	[–0.22, –0.13]	< .001
Suboptimal level				0.52	[0.47, 0.57]	< .001
Morality	**–0.05**	**[–0.09, –0.02]**	**.003**	**–0.07**	**[–0.10, –0.04]**	**< .001**
Social connectedness						
Low level	0.19	[0.16, 0.23]	< .001	–0.01	[–0.06, 0.04]	.802
Suboptimal level				0.26	[0.22, 0.31]	< .001
Morality	0.01	[–0.02, 0.05]	.455	*–0.01*	*[–0.04, 0.02]*	*.47*
Meaning						
Low level	0.12	[0.08, 0.15]	< .001	–0.15	[–0.20, –0.11]	< .001
Suboptimal level				0.41	[0.36, 0.45]	< .001
Morality	0.01	[–0.03, 0.05]	.49	0.05	[0.02, 0.09]	.003
Social status						
Low level	0.07	[0.03, 0.10]	< .001	–0.17	[–0.22, –0.12]	< .001
Suboptimal level				0.36	[0.31, 0.41]	< .001
Morality	**0.08**	**[0.04, 0.12]**	**< .001**	**0.08**	**[0.05, 0.12]**	**< .001**

Note. Effects of morality that remained statistically detectable at a $p < .05$ threshold are shown in **boldface**. Effects of morality that were no longer statistically detectable after controlling for perceived deficiencies (but were statistically detectable in Table 3) are shown in *italics*. Effects of morality that remained statistically detectable at a $p < .05$ threshold are underlined. This analysis was preregistered in Study 2b but not Study 2a. Study 2a did not measure suboptimal levels. See Table 3 for effects of morality without controlling for perceived deficiencies.

4.2. Individual Differences in Predictors of Change Goals

As shown in Table 1 and Figure 3, there was a substantial amount of slope variance (i.e., between-person variability in within-person slopes) for the belief variables. Considering that people likely value different outcomes (e.g., happiness vs. meaning) to different extents, can individual differences in predictors of change goals be explained by people's explicit valuation of various outcomes? Here, we consider the extent to which people value a given outcome

compared to other people (between-person analyses) and the extent to which people value a given outcome compared to other outcomes (profile analyses).

Between-Person Analyses

First, the extent to which a person's personality change goals are tied with beliefs about consequences for a given outcome might depend on how much they want to improve that outcome compared to other people. For example, Morgan's personality change goals may be more strongly predicted by her beliefs about happiness consequences than are Logan's personality change goals. This could be because Morgan is more interested in improving her happiness than is Logan.

To measure participants' desire to change each of the outcomes, we used the same item format and response scale as for the input trait change goals, but with appropriate wording modifications (e.g., "To what extent would you like to change the extent to which you are achieving the various goals you have?"; –3 = *I want to achieve many fewer of my goals*, 0 = *I don't want to change how many of my goals I am achieving*, +3 = *I want to achieve many more of my goals*). On average, participants wanted to see the greatest improvements in their goal attainment (Study 2a: $M = 2.12$, $SD = 1.12$; Study 2b: $M = 2.05$, $SD = 1.02$), followed by happiness (Study 2a: $M = 2.07$, $SD = 1.08$; Study 2b: $M = 1.71$, $SD = 1.04$), meaning (Study 2a: $M = 1.72$, $SD = 1.16$; Study 2b: $M = 1.46$, $SD = 1.13$), connectedness (Study 2a: $M = 1.70$, $SD = 1.12$; Study 2b: $M = 1.45$, $SD = 1.10$), status (Study 2a: $M = 1.51$, $SD = 1.16$; Study 2b: $M = 1.26$, $SD = 1.14$), and morality (Study 2a: $M = 0.80$, $SD = 1.09$; Study 2b: $M = 0.39$, $SD = 0.84$).

To test whether outcome change goals moderated the within-person associations between beliefs about the respective outcome and personality change goals, we fit six linear mixed-effects models (one outcome per model), in which we predicted personality change goals from beliefs

about consequences for that outcome, the extent to which they wanted to improve that outcome, and the cross-level interaction. As shown in Table S8, the interaction effect was significant and positive for all outcomes except for meaning in life in Study 2a and social status in Study 2b. This means that the within-person effects of beliefs about consequences for morality, happiness, goal attainment, social connectedness, social status (Study 2a), and meaning in life (Study 2b) were more positive for participants who had a greater desire to improve these respective outcomes. In other words, people are especially interested in improving the traits that they believe will improve a given outcome (e.g., happiness) if they are more interested in improving that outcome (compared with other people).

Table S8

Predicting Change Goals from Beliefs About Consequences, Outcome Change Goals, and Their Interaction (Studies 2a and 2b)

	Beliefs About Consequences		Outcome Change Goals		Cross-Level Interaction			
Predictor	β	95% CI	β	95% CI	β	95% CI	*p*	R^2
Study 2a								
Morality	–0.11	[–0.14, –0.07]	0.32	[0.27, 0.37]	**0.08**	**[0.05, 0.11]**	**< .001**	.23
Social status	0.27	[0.23, 0.30]	0.28	[0.23, 0.33]	**0.07**	**[0.03, 0.11]**	**< .001**	.11
Happiness	0.39	[0.35, 0.42]	0.31	[0.25, 0.37]	**0.07**	**[0.04, 0.11]**	**< .001**	.08
Goal attainment	0.46	[0.42, 0.49]	0.4	[0.34, 0.46]	**0.05**	**[0.01, 0.08]**	**.011**	.07
Connectedness	0.18	[0.15, 0.22]	0.32	[0.26, 0.38]	**0.05**	**[0.02, 0.09]**	**.005**	.02
Meaning	0.29	[0.25, 0.33]	0.3	[0.24, 0.35]	0.03	[–0.01, 0.07]	.164	.01
Study 2b								
Connectedness	0.23	[0.19, 0.26]	0.22	[0.17, 0.27]	**0.07**	**[0.04, 0.10]**	**< .001**	.13
Happiness	0.40	[0.36, 0.43]	0.22	[0.17, 0.27]	**0.08**	**[0.04, 0.11]**	**< .001**	.12
Morality	–0.13	[–0.17, –0.10]	0.16	[0.12, 0.21]	**0.05**	**[0.02, 0.08]**	**.001**	.09
Meaning	0.32	[0.29, 0.35]	0.18	[0.13, 0.22]	**0.05**	**[0.01, 0.08]**	**.008**	.05
Goal attainment	0.48	[0.45, 0.51]	0.28	[0.23, 0.33]	**0.04**	**[0.00, 0.07]**	**.025**	.02
Social status	0.30	[0.27, 0.34]	0.21	[0.17, 0.25]	0.01	[–0.03, 0.04]	.733	–.01

Note. Key interaction effects that were statistically detectable at an uncorrected $p < .05$ threshold are shown in **boldface**. R^2 = proportion of slope variance explained by adding the cross-level interaction. All *p* values for beliefs about consequences and outcome change goals were < .001. Predictors are ordered from the largest to smallest cross-level interactions (based on the R^2).

Profile Analyses

Individual differences in predictors of personality change goals can also be examined using a profile approach (i.e., the relative importance of each of the predictors within a given person). For example, Jane's personality change goals may be more strongly related to her beliefs about happiness consequences than her beliefs about morality consequences, whereas John's personality change goals may be more strongly related to his beliefs about morality consequences than his beliefs about happiness consequences. Note that in the Study 2a preregistration, we planned to include perceived deficiencies in this profile. However, we deviated from this plan as perceived deficiencies was assessed on a different metric than beliefs about consequences for various outcomes, which meant that the unstandardized regression coefficients were not directly comparable. We therefore only included beliefs about consequences for various outcomes in the profile (a decision that we preregistered in Study 2b).

On average, participants generally placed greater positive weight on beliefs about consequences for goal attainment (Study 2a: $b = 0.43$; Study 2b: $b = 0.45$), followed by happiness (Study 2a: $b = 0.41$; Study 2b: $b = 0.41$), meaning (Study 2a: $b = 0.33$; Study 2b: $b = 0.36$), status (Study 2a: $b = 0.31$; Study 2b: $b = 0.33$), social connectedness (Study 2a: $b = 0.20$; Study 2b: $b = 0.24$), and morality (Study 2a: $b = –0.13$; Study 2b: $b = –0.15$). To quantify the extent to which people adhered to the normative (i.e., average) profile, we extracted the individual slopes for each person, rank-ordered them within-person, and computed the percentage of participants who showed an exact match with the average rank-ordering. Given that the unstandardized fixed effects for goal attainment vs. happiness and social status vs. connectedness (respectively) were almost identical in both studies, we felt this preregistered exact match metric was probably too stringent. Thus, we computed the percentage who matched

a more lenient (non-preregistered) metric, in which a profile counted as a match if goal attainment and happiness were the top two most positive predictors, followed by meaning and social status (ranks = 3 or 4), social connectedness (rank = 5), perceived deficiencies (rank = 6), and morality (rank = 7).

Even under this more lenient metric, this "average" profile described less than a quarter of participants (Study 2a: 23.43%; Study 2b: 21.26%). In fact, there were 81 (Study 2a) or 71 (Study 2b) distinct profiles of rank-orderings (see Panel A of Figures S5–S6). Notably, however, beliefs about morality consequences was among the bottom two predictors of change goals for 99% of participants in both studies, and not a single participant's personality change goals were most positively predicted by their beliefs about consequences for morality.

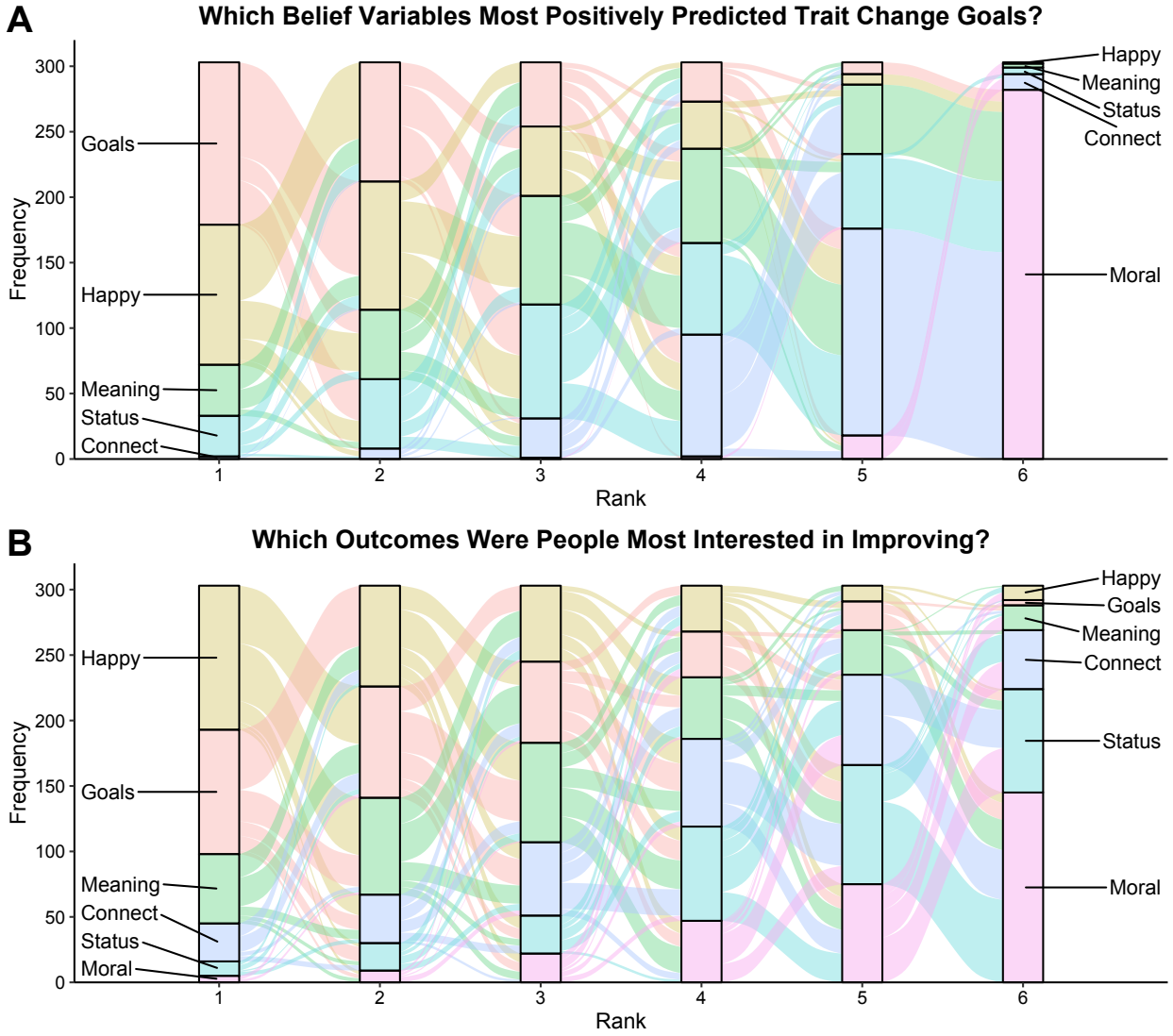
Do these distinct profiles of predictors (see Panel A, Figures S5–S6) correspond to people's explicit rankings of which outcomes they most want to improve (see Panel B, Figures S5–S6)? For the rank-ordering question, we asked participants, "Which outcomes are you most interested in improving? Please rank them from 1 (most interested in improving) to 6 (least interested in improving)." On average, participants most wanted to improve their happiness (Study 2a: M_{rank} = 2.32; Study 2b: M_{rank} = 2.37) and goal attainment (Study 2a: M_{rank} = 2.39; Study 2b: M_{rank} = 2.36), followed by meaning (Study 2a: M_{rank} = 2.97; Study 2b: M_{rank} = 3.09), connectedness (Study 2a: M_{rank} = 3.81; Study 2b: M_{rank} = 3.98), status (Study 2a: M_{rank} = 4.48; Study 2b: M_{rank} = 4.39), and morality (Study 2a: M_{rank} = 5.02; Study 2b: M_{rank} = 4.79). We did not previously preregister a plan for dealing with ties in ranks. However, given that there was no statistically significant difference in the ranks for happiness and goal attainment in a non-preregistered paired-samples *t*-test in both studies (*p*s > .567), we decided to count these as being tied in first place. Non-preregistered paired-samples *t*-tests also showed that in both studies,

happiness and goal attainment were ranked higher than meaning (*p*s < .001), which was ranked higher than social connectedness (*p*s < .001), which was ranked higher than social status (*p*s < .001), which was ranked higher than morality (*p*s < .003). We therefore defined the "average" profile as Rank 1 = either happiness or goal attainment, Rank 2 = either happiness or goal attainment, Rank 3 = meaning, Rank 4 = connectedness, Rank 5 = status, and Rank 6 = morality. Only 2.64% (Study 2a) and 3.99% (Study 2b) of participants adhered to this average profile; indeed, there were 160 (Study 2a) and 174 (Study 2b) distinct profiles of rank-orderings (see Panel B of Figures S5–S6). We examined the profile correlations between this rank-ordering measure and within-person predictors of change goals.

To examine the correlations between profiles of within-person effects and explicit rankings of outcome change goals, we computed overall and distinctive profile correlations using the *multicon* package (Sherman & Serfass, 2015). Overall correlations include agreement based on normativity—for example, agreement may be high not because an individual's explicit rankings are uniquely associated with their within-person slopes, but because on average, most people tend to rank happiness as being more important than meaning, and also tend to have stronger within-person associations between personality change goals and happiness beliefs than meaning beliefs. Distinctive correlations capture the extent of unique agreement once such average effects are removed (Furr, 2008). In both cases, the *p* values for both the overall and distinctive correlations reflect whether agreement reflects something unique about the individual (vs. rank-orderings for an average person).

Figure S5

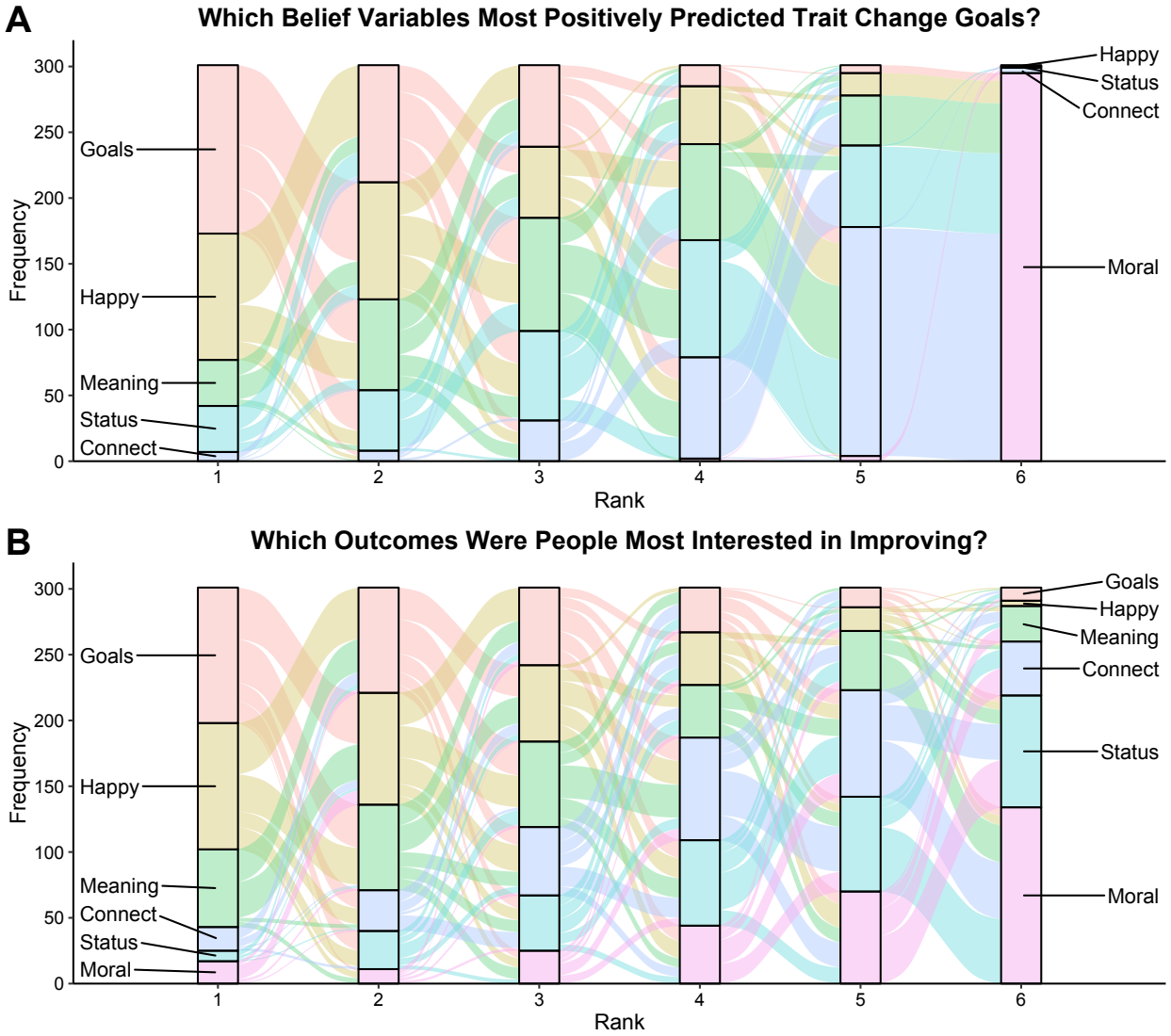
Rank-Orderings of (a) Reasons for Trait Change Goals and (b) Interest in Improving Each Outcome (Study 2a)



Note. Panel A shows the rank-orderings of the within-person effects of perceived deficiencies and beliefs about consequences for each outcome on trait change goals, from 1 (most positive predictor) to 6 (least positive predictor). Panel B shows participants' explicit rankings of which outcomes they are most interested in improving, from 1 (most interested) to 6 (least interested). Happy = Day-to-day happiness, Goals = Goal attainment, Meaning = Meaning in life, Connect = Social connectedness, Status = Social status, Moral = Morality. These alluvial diagrams visualize the flow of data across rank-orderings. For example, beliefs about goal attainment consequences was the most positive predictor for 124 participants. Of these participants, beliefs about happiness consequences was the second most positive predictor for 72 participants.

Figure S6

Rank-Orderings of (a) Reasons for Trait Change Goals and (b) Interest in Improving Each Outcome (Study 2b)



Note. Panel A shows the rank-orderings of the within-person effects of perceived deficiencies and beliefs about consequences for each outcome on trait change goals, from 1 (most positive predictor) to 6 (least positive predictor). Panel B shows participants' explicit rankings of which outcomes they are most interested in improving, from 1 (most interested) to 6 (least interested). Happy = Day-to-day happiness, Goals = Goal attainment, Meaning = Meaning in life, Connect = Social connectedness, Status = Social status, Moral = Morality. These alluvial diagrams visualize the flow of data across rank-orderings. For example, beliefs about goal attainment consequences was the most positive predictor for 128 participants. Of these participants, beliefs about happiness consequences was the second most positive predictor for 60 participants.

In Study 2a, there was a substantial positive overall profile correlation ($r = .56$, $p = .002$) and a small distinctive profile correlation (Study 2a: $r = .07$, $p = .018$). In Study 2b, the overall ($r = .52$, $p = .148$) and distinctive ($r = .03$, $p = .187$) profile correlations were of a similar magnitude as in Study 2a, but were not statistically detectable. Thus, people's explicit rankings of these change goals correspond to the relative ordering of the different reasons, but this was either largely (Study 2a) or entirely (Study 2b) driven by average rankings of these predictors and outcomes across people.

Conclusion

These individual difference findings suggest a further nuance to the Personal Fulfillment Hypothesis. Although most people are more interested in improving the traits that would increase their well-being to a greater extent, beliefs about consequences for different dimensions of well-being more strongly predict change goals to the extent that people explicitly value these dimensions compared to other people. Moreover, although day-to-day happiness and goal attainment were the strongest predictors on average, beliefs about consequences for meaning, connectedness, or social status were the strongest predictors for some people (see Figures S5–S6). Importantly, however, morality was never the most important predictor for any participant. This suggests that although the Personal Fulfillment Hypothesis appears to be generally supported, the specific ways in which it manifests differs across people.

5. Supplemental Study

The primary goal of this Supplemental Study was to address a reviewer's concern that the perceived tradeoff between morality and personal fulfillment may have been due to the mention of "personal sacrifice" in the definition of morality. To address this concern, we investigated whether participants still perceived a tradeoff between morality and personal fulfillment when

the definition of morality no longer mentioned personal sacrifice. We also investigated whether happiness and goal attainment continued to predict change goals after controlling for other potential explanations—difficulty, controllability, and discrepancy between current and optimal levels of traits.

To do so, we conducted a partial replication and extension of Study 2a. Unlike Study 2a, participants in the Supplemental Study (a) saw a definition of morality that did not mention self-sacrifice, (b) only reported their beliefs about consequences for happiness and goal attainment (as these were the only two dimensions of personal fulfillment that showed tradeoffs with morality in Study 2a), and (c) also reported on the discrepancy between their current and optimal levels of each trait (as an additional indicator of perceived deficiencies).

We used the Supplemental Study data in two ways. First, we computed crowdsourced norms for each trait's morality, difficulty, and controllability by averaging Supplemental Study participants' ratings of these variables. We then re-ran certain analyses in Study 2a, replacing the confounded morality variable and adding the new difficulty and controllability variables. A limitation of the first strategy is that the same morality norms are applied equally across the participants in Study 2a. However, in reality, people differ in how much they think various improvements will affect their morality, and as can be seen in Table S11, effect sizes using norms are not directly comparable with effect sizes using person-specific scores. As a second strategy, we therefore conduct person-specific analyses using the Supplemental Study data.

Method

Data collection procedures were approved by the Research Ethics Committee at [Institution Name Masked] (REC ID: REC771-11102024). We preregistered stopping rules, exclusions, and analysis plans at

https://osf.io/ux5en/?view_only=1aeb7dbb437841d6a915d8536d6742c4. All study materials and the data and analysis code are available at https://osf.io/xq6sf/?view_only=839853c333ec4afbb546bb8b12e77c18. We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

Participants

We aimed to recruit 100 UK-based participants via Prolific Academic. Participants were compensated with £4.00 for completing a 20–30-minute survey. 115 participants provided consent and started the survey. As preregistered, we excluded 4 participants who failed either of the two comprehension checks, as well as one participant who did not submit the survey. After applying these exclusions, the final sample comprised 110 participants (44 men, 66 women) who ranged in age from 19 to 75 years (M = 42.73, SD = 13.12). Participants identified as Asian/Asian British (n = 6), White/White British (n = 90), Black/Black British (n = 9), Hispanic or Latino (n = 1), or Other or Multiple (n = 4).

Procedure

Participants viewed the definitions of 3 outcomes (day-to-day happiness, goal attainment, and morality; see Appendix B), before completing a comprehension check in which they were required to correctly match up each outcome with its definition. They then viewed the definitions of 22 input traits (see Appendix B), before completing a comprehension check in which they were required to correctly match up a subset of 7 or 8 traits with their definitions. As in Study 2a, participants who failed either comprehension check were automatically screened out of the survey.

Participants who passed both comprehension checks proceeded to the rest of the survey. Participants reported (a) their beliefs about how improving each of the 22 input traits would

change each of the 3 outcomes, (b) how difficult and within their control changing each of the 22 input traits would be, and (c) their perceived deficiencies and change goals for the input traits (described below). We counterbalanced whether participants reported their perceived deficiencies and change goals first or last.

Measures

Input Trait Change Goals. Participants completed the same measure of change goals as in Study 2. As in Study 2, scores for the negative traits (Anxiety, Depression, and Exploitativeness) were reverse-scored, such that higher scores on all change goals indicate a greater desire to *improve* each respective trait.

Perceived Deficiencies. Participants reported their current levels and suboptimal levels using the same measures as in Study 2b. As in Studies 1–2, for current levels, we reverse-scored all items except for the three negative traits (Anxiety, Depression, and Exploitativeness), such that higher scores reflect greater perceived deficiencies. As in Study 2b, for suboptimal levels, positive scores correspond to a person believing that they are currently below their optimal level of a trait, whereas negative scores correspond to a person believing that they are currently above their optimal level of a trait.

Beliefs about Consequences of Personality Improvements. As in Studies 1–2, participants imagined a change that would make a noticeable difference in their daily. They then reported their beliefs about how improving each of the 22 input traits (e.g., “If I became more [efficient], I would:”) would change each their day-to-day happiness, goal attainment, and morality, using the same response scale as in the previous studies. For the input traits of anxiety, depression, and exploitativeness, as in Study 2, the items were phrased in terms of becoming *less*

anxious, depressed, and exploitative (respectively). Participants answered 66 items (22 input traits × 3 outcomes).

Difficulty. Participants reported on how difficult it would be to improve on each input trait, using one item per trait (e.g., "How difficult would it be for you to become more [sociable]?"; 0 = *Not difficult at all*, 10 = *Extremely difficult*). For anxiety, depression, and exploitativeness, participants were asked about the difficulty of reducing these traits.

Controllability. Participants reported on the extent to which it would be within their control to improve on each input trait, using one item per trait (e.g., "To what extent do you think that becoming more [sociable] is within your control?"; 0 = *Not at all within my control*, 10 = *Completely within my control*). For anxiety, depression, and exploitativeness, participants were asked about the controllability of reducing these traits.

Results

Descriptive Statistics

Descriptive statistics and intercorrelations are shown in Tables S8 and S9.

Table S9

Descriptive Statistics and Correlation Matrix for Perceived Deficiencies, Beliefs About Consequences of Personality Improvements, Difficulty, and Controllability (Supplemental Study)

				Intercorrelations							
Variable	*M*	SD_{WP}	SD_{BP}	1.	2.	3.	4.	5.	6.	7.	8.
1. Change goals	0.98	0.65	1		.24	.61	.56	.59	.44	–.26	.10
2. Low levels	3.4	0.94	2.17	.40		.40	–.23	–.15	–.25	.36	–.22
3. Suboptimal levels	1.28	1.13	2.33	.52	.76		.26	.25	.11	–.05	.35
4. Happiness beliefs	0.96	0.62	1.06	.40	.24	.37		.79	.70	–.34	.23
5. Goal beliefs	0.91	0.6	1.14	.43	.25	.34	.40		.74	–.21	.11
6. Morality beliefs	0.55	0.56	0.87	–.01	–.10	–.06	.01	–.13		–.11	.08
7. Difficulty	4.36	1.48	2.42	.33	.51	.41	.20	.26	–.14		–.34
8. Controllability	7.15	1.49	1.98	–.18	–.34	–.31	–.22	–.15	.19	–.45	

Note. Correlations below the diagonal are within-person; correlations above the diagonal are between-person.

Table S10

Means (SDs) for Change Goals and Predictors of Change Goals (Supplemental Study)

				Beliefs About Consequences For…				
Input Trait	Change goals	Low level	Subopt. level	Happy	Goals	Morality	Difficulty	Control
Energy	1.66	4.89	2.83	1.55	1.77	0.29	5.43	6.58
	(1.10)	(2.03)	(2.50)	(1.06)	(1.11)	(0.77)	(2.49)	(2.23)
(–) Anxiety	1.58	5.88	4.30	2.03	1.55	0.22	6.27	4.68
	(1.42)	(2.70)	(3.28)	(1.13)	(1.23)	(0.81)	(2.75)	(2.47)
Creativity	1.45	4.15	1.91	1.41	1.34	0.22	4.73	6.35
	(1.17)	(2.40)	(2.20)	(1.14)	(1.19)	(0.70)	(2.68)	(2.67)
Intelligence	1.44	3.32	1.56	0.87	1.54	0.55	6.12	5.77
	(1.17)	(1.57)	(2.04)	(1.14)	(1.15)	(1.04)	(2.55)	(2.89)
(–) Depression	1.38	4.08	3.04	2.11	1.61	0.27	6.07	4.61
	(1.45)	(2.88)	(3.09)	(1.12)	(1.25)	(0.82)	(2.62)	(2.66)
Self-Control	1.30	4.13	2.25	0.89	1.46	0.50	4.96	7.59
	(1.15)	(2.13)	(2.57)	(1.22)	(1.18)	(0.89)	(2.54)	(2.16)
Organization	1.30	3.30	1.42	1.08	1.66	0.28	3.95	8.15
	(1.09)	(1.93)	(2.20)	(1.08)	(1.05)	(0.76)	(2.53)	(1.86)
Emotional Stability	1.27	3.78	1.91	1.49	1.02	0.51	5.16	5.94
	(1.24)	(2.29)	(2.75)	(1.15)	(1.10)	(0.90)	(2.58)	(2.48)
Sociability	1.25	4.95	1.98	1.41	0.80	0.26	5.46	7.06
	(1.14)	(2.55)	(2.34)	(1.17)	(1.30)	(0.93)	(2.70)	(2.13)
Efficiency	1.25	3.16	1.48	1.08	1.83	0.27	4.17	7.89
	(1.11)	(1.98)	(2.41)	(1.14)	(1.10)	(0.83)	(2.25)	(1.72)
Curiosity	1.18	2.93	1.01	1.03	1.18	0.44	3.83	7.39
	(1.07)	(1.98)	(1.93)	(1.04)	(1.17)	(0.93)	(2.46)	(2.29)
Assertiveness	1.04	4.69	1.46	0.58	1.36	0.17	5.50	6.98
	(1.18)	(2.11)	(2.31)	(1.16)	(1.09)	(0.96)	(2.35)	(2.18)
Aesthetic Sensitivity	0.82	2.93	0.71	1.40	0.35	0.42	3.23	7.72
	(0.91)	(2.28)	(1.80)	(1.14)	(1.04)	(0.89)	(2.54)	(2.17)
Principledness	0.71	2.75	0.41	0.50	0.65	1.02	3.82	8.07
	(1.03)	(1.74)	(1.95)	(1.03)	(1.28)	(1.11)	(2.46)	(1.91)
Responsibility	0.68	2.08	–0.23	0.33	0.92	0.77	3.70	8.09
	(1.08)	(1.51)	(1.97)	(1.08)	(1.13)	(1.06)	(2.43)	(1.73)
Compassion	0.67	2.66	0.46	0.71	0.24	1.05	3.38	7.64
	(1.01)	(1.66)	(1.73)	(1.04)	(0.90)	(1.20)	(2.44)	(2.06)
Honesty	0.56	2.16	0.28	0.49	0.21	1.09	2.93	8.35
	(0.93)	(1.58)	(1.89)	(0.96)	(0.86)	(1.10)	(2.80)	(1.90)
Fairness	0.51	2.14	0.46	0.34	0.11	1.02	3.20	8.09
	(0.85)	(1.28)	(1.55)	(0.88)	(0.94)	(1.12)	(2.56)	(1.87)
Humility	0.41	2.89	–0.01	0.56	0.10	0.69	3.25	7.60
	(0.94)	(1.82)	(2.56)	(0.98)	(0.92)	(1.04)	(2.61)	(2.25)
Trust	0.38	3.87	0.97	0.50	0.22	0.48	5.16	6.69
	(1.00)	(2.35)	(2.28)	(1.09)	(1.06)	(0.95)	(2.62)	(2.17)
(–) Exploitativeness	0.35	1.98	0.14	0.44	–0.05	1.05	2.49	8.16
	(1.09)	(2.18)	(2.40)	(1.08)	(0.97)	(1.37)	(2.66)	(2.22)
Loyalty	0.33	2.09	–0.21	0.30	0.07	0.62	3.08	7.86
	(0.76)	(1.71)	(1.76)	(0.81)	(0.96)	(1.02)	(2.56)	(1.93)

Note. All means were detectably different from 0 at an FDR-corrected $p < .05$ level, except for the effects of exploitativeness, loyalty, fairness, and humility on goal attainment, and the effect of assertiveness on morality. Control = controllability.

Do People Perceive Tradeoffs Between Morality and Personal Fulfillment?

The primary goal of this supplemental study was to address a reviewer's concern that the perceived tradeoff between morality with personal fulfillment may have been due to the mention of "personal sacrifice" in the definition of morality. We therefore investigated whether participants still perceived a tradeoff between morality and personal fulfillment when the definition of morality no longer mentioned personal sacrifice. We used the Supplemental Study data in two ways.

Replacing Study 2a Participants' Moral Belief Ratings with Unconfounded Norms. First, we computed the mean of Supplemental Study participants' beliefs about how various changes would increase or decrease their morality. This yielded crowdsourced norms for the average beliefs about how each of 22 traits would impact a person's morality (when the definition of morality does not include self-sacrifice). We re-ran the key analyses for Study 2a using these unconfounded crowdsourced norms instead of the Study 2a participants' own (person-specific) ratings of how much they believed that each trait change would affect morality.

As shown in Table S11, participants in Study 2a believed that less morally relevant improvements (according to participants in the Supplemental Study) would increase goal attainment and happiness to a greater extent than would more morally relevant improvements. Unexpectedly, tradeoffs also emerged for connectedness, meaning, and social status. Moreover, the effect sizes for the analyses that involved the unconfounded morality norms were much larger than for the original analyses that involved person-specific estimates.

Table S11

Predicting Beliefs About Consequences for Personal Fulfillment from Beliefs About Consequences for Morality (Using Unconfounded Moral Norms) for Study 2a

	Original Person-Specific Morality Beliefs from Study 2a			Unconfounded Morality Norms from the Supplemental Study			Potentially Confounded Morality Norms from Study 2a (Not Preregistered)		
Outcome	β/*b*	95% CI	*p*	β/*b*	95% CI	*p*	β/*b*	95% CI	*p*
Goal attainment	–0.18	[–0.21, –0.14]	< .001	–0.41	[–0.44, –0.38]	< .001	–0.40	[–0.43, –0.37]	< .001
Happiness	–0.08	[–0.12, –0.04]	< .001	–0.30	[–0.33, –0.26]	< .001	–0.30	[–0.33, –0.26]	< .001
Connectedness	–0.01	[–0.05, 0.03]	.726	–0.18	[–0.21, –0.14]	< .001	–0.17	[–0.21, –0.14]	< .001
Meaning	–0.00	[–0.04, 0.04]	.910	–0.16	[–0.19, –0.12]	< .001	–0.18	[–0.22, –0.15]	< .001
Social status	0.07	[0.03, 0.11]	< .001	–0.05	[–0.09, –0.01]	.02	–0.06	[–0.10, –0.02]	.005

Note. Unconfounded Morality Norms = Supplemental Study participants' mean beliefs about how improving various traits would improve morality (in which the definition of morality did not mention self-sacrifice). Potentially Confounded Morality Norms = Study 2a participants' mean beliefs about how improving various traits would improve morality (in which the definition of morality mentioned self-sacrifice).

To investigate whether these larger effect sizes may have been due to the aggregation of the morality norms, we conducted exploratory (non-preregistered) analyses in which we ran the same analyses using aggregated morality ratings from Study 2a. As shown in Table S11, the effect size estimates were almost identical for the analyses that used aggregated morality ratings (whether they came from Study 2a, in which the definition mentioned self-sacrifice, or the Supplemental Study, in which the definition did not mention self-sacrifice). This is not that surprising, given that average ratings of which traits were more vs. less morally relevant correlated .95 ($p < .001$) between Study 2a and the Supplemental Study.

Person-Specific Analyses with Supplemental Study Data. We also conducted a limited replication using the person-specific data from the Supplemental Study. This was a limited replication in that we only collected new data for the outcomes of day-to-day happiness and goal attainment (i.e., cutting meaning in life, connectedness, and status).

It is plausible that people perceive diminishing marginal returns, such that people imagine that the same amount of improvement in a trait would yield greater happiness benefits (for example) when they are further away from the optimal level on that trait. We therefore ran two sets of analyses, either controlling or not controlling for current levels and suboptimal levels. In both sets of analyses, participants in the Supplemental Study perceived a tradeoff between the improvements that they considered to be more morally relevant and the improvements that they thought would help them achieve their goals to a greater extent (as in Study 2a; see Table S12). However, no such tradeoff emerged for happiness for these person-specific analyses (unlike Study 2a).

Table S12

Predicting Beliefs About Consequences for Happiness and Goal Attainment from Perceived Deficiencies and Beliefs About Consequences for Morality (Supplemental Study)

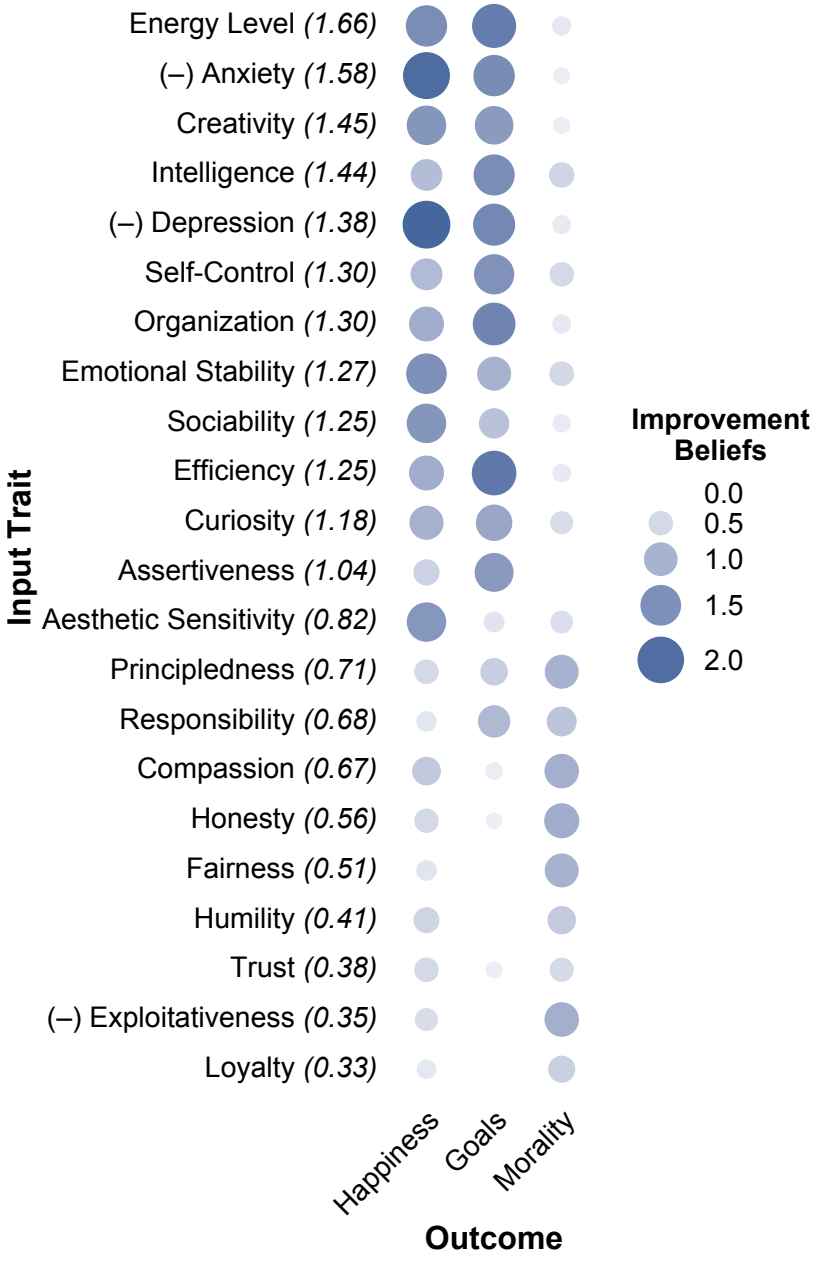
	Not Controlling for Perceived Deficiencies (Not Preregistered)			Controlling for Perceived Deficiencies		
Outcome	β	95% CI	*p*	β	95% CI	*p*
Goal attainment						
Low level				–0.06	[–0.14, 0.03]	.176
Suboptimal level				0.36	[0.28, 0.44]	< .001
Morality	–0.18	[–0.26, –0.11]	< .001	–0.16	[–0.22, –0.09]	< .001
Happiness						
Low level				–0.12	[–0.21, –0.03]	.009
Suboptimal level				0.45	[0.37, 0.54]	< .001
Morality	–0.04	[–0.12, 0.03]	.238	0.00	[–0.07, 0.07]	.998

Note. Not Controlling for Perceived Deficiencies was not preregistered, but we included it as we thought it would be informative to facilitate comparison with Table S11.

Does this mean that participants in the Supplemental Study did not perceive a tradeoff between happiness and morality? A non-preregistered analyses showed that the by-traits correlation between morality and happiness was almost identical between Study 2a ($r = –.66$, $p < .001$) and the Supplemental Study ($r = –.68$, $p < .001$). In other words, the Supplemental Study still showed that the traits that participants generally thought were more morally relevant were the traits that participants expected would increase happiness to a lesser extent (see Figure S7).

Figure S7

Mean Beliefs about the Consequences of Each Trait Improvement on All Outcomes in the Supplemental Study



Note. The size of each circle represents the average (mean) extent to which people believe that improving each input trait (*x*-axis) would improve each outcome (*y*-axis). Input traits are ordered from the traits that participants most wanted to improve (energy level) to the ones that they least wanted to improve (loyalty). All means were detectably different from 0 at an FDR-corrected $p < .05$ level, except for the effects of exploitativeness, loyalty, fairness, and humility on goal attainment, and the effect of assertiveness on morality (not visualized). Mean change goals for each input trait are shown in parentheses (where the response scale corresponded to 0 = no change desired, 1 = slightly more, 2 = more, 3 = much more). See Table S10 for means and standard deviations.

Predicting Change Goals: Robustness Check and Extensions

Robustness Check for Study 2a. For completeness, we also conducted a robustness check to examine whether the relative importance of the various predictors of change goals changed when self-sacrifice was removed from the definition of morality. To do so, we re-ran the key models from Study 2a (reported in Table 2) after replacing Study 2 participants' moral belief ratings with crowdsourced norms derived from the Supplemental Study.

First, we fit a new single-predictor model in which we predicted Study 2a participants' change goals from the new, unconfounded morality consequences variable (based on crowdsourced estimates in the Supplemental Study). Participants in Study 2a were less interested in improving traits that participants in the Supplemental Study generally considered to be more morally relevant, $\beta = –0.31$, 95% CI [–0.34, –0.27], $p < .001$.

Next, we fit a seven-predictor model in which we predicted Study 2a participants' change goals change goals from six individual-level predictors reported by Study 2a participants (perceived deficiencies and beliefs about consequences for happiness, goal attainment, meaning, connectedness, and social status) and the new, unconfounded morality consequences variable (based on crowdsourced estimates in the Supplemental Study). As shown in Table S13, participants in Study 2a continued to be less interested in improving traits that participants in the Supplemental Study generally considered to be more morally relevant (an effect that was not uniquely detectable in the multiple-predictor model that used person-specific moral relevance estimates in Study 2a; see Table 2). Apart from that, the effect sizes for the other predictors barely changed (maximum change = |0.02|).

Table S13

Multiple-Predictor Models Predicting Change Goals from Perceived Deficiencies, Beliefs About Consequences for Morality and Well-Being Outcomes, Difficulty, and Controllability (Study 2a)

	Robustness Check			Extension		
Predictor	β	95% CI	*p*	β	95% CI	*p*
Goal attainment	0.28	[0.25, 0.31]	< .001	0.27	[0.24, 0.30]	< .001
Happiness	0.16	[0.13, 0.20]	< .001	0.17	[0.14, 0.20]	< .001
Low levels	0.16	[0.12, 0.19]	< .001	0.15	[0.12, 0.18]	< .001
Social status	0.12	[0.10, 0.15]	< .001	0.12	[0.10, 0.15]	< .001
Meaning	0.09	[0.06, 0.12]	< .001	0.09	[0.07, 0.12]	< .001
Connectedness	–0.01	[–0.04, 0.01]	.229	–0.02	[–0.04, 0.01]	.188
Morality	–0.07	[–0.10, –0.05]	< .001	–0.06	[–0.09, –0.04]	< .001
Difficulty				0.07	[0.02, 0.11]	.006
Controllability				0.06	[0.01, 0.11]	.016

Note. τ^2 = slope variance. Morality, difficulty, and controllability were based on norms from the Supplemental Study (and were therefore not person-specific). The other predictors were person-specific scores provided by Study 2a participants.

Extensions: Difficulty, Controllability, and Suboptimality. Reviewers also raised two additional possible explanations for why people may be less interested in being more moral: 1) People might think it is harder to improve moral traits, and 2) People might already be satisfied with how moral they are. To assess beliefs about changeability, we included measures of perceived difficulty and controllability. To assess satisfaction with current levels of morality, we included an additional operationalization of perceived deficiencies—the discrepancy between optimal and current levels—and adjusted for both operationalizations of perceived deficiencies (low levels and suboptimal levels) when estimating the effects of various predictors of change goals.

Difficulty and Controllability. First, we tested whether people think that moral traits are less changeable. In fact, participants in the Supplemental Study thought that more morally relevant improvements would be less difficult (β = –0.15, 95% CI [–0.22, –0.09], $p < .001$) and more within their control (β = 0.18, 95% CI [0.12, 0.24], $p < .001$).

Single-predictor models showed that participants in Study 2a were more interested in improving traits that participants in the Supplemental Study considered to be *less* controllable (β = –0.22, 95% CI [–0.26, –0.19], $p < .001$) and *more* difficult to change (β = 0.30, 95% CI [0.26, 0.33], $p < .001$). These effects also emerged in person-specific analyses predicting Supplemental Study participants' change goals from their own ratings of traits' difficulty and changeability (see Table S14).

Importantly, however, when difficulty and controllability norms from Supplemental Study participants were added to the model predicting Study 2a participants' change goals from perceived deficiencies and beliefs about consequences for happiness, goal attainment, meaning, connectedness, and social status, the estimates for the other predictors barely changed (though the direction of the controllability predictor flipped; see Table S13). That is, controlling for difficulty and controllability, participants in Study 2a were still more interested in improving traits that would better improve their well-being (especially their goal attainment and happiness), and less interested in improving more morally relevant traits.

In person-specific analyses in the Supplemental Study dataset (described further below), goal attainment and happiness were also the most important predictors, even controlling for difficulty and controllability (see Table S14). Indeed, difficulty and controllability only explained an additional 2% of variance in change goals (over and above perceived deficiencies, morality, and beliefs about consequences for goal attainment and happiness).

Table S14

Predicting Change Goals from Perceived Deficiencies, Beliefs About Consequences for Morality and Well-Being Outcomes, Difficulty, and Controllability (Supplemental Study)

	Single-Predictor Models				Multiple-Predictor Models			
Predictor	β	95% CI	*p*	τ^2	β	95% CI	*p*	R^2
Step 1								.36
Low level					0.06	[–0.03, 0.15]	.176	
Suboptimal level					0.45	[0.38, 0.53]	< .001	
Step 2								.38
Low level					0.05	[–0.03, 0.14]	.224	
Suboptimal level					0.45	[0.38, 0.53]	< .001	
Morality					0.00	[–0.04, 0.05]	.92	
Step 3								.47
Low level					0.07	[–0.00, 0.15]	.064	
Suboptimal level					0.29	[0.22, 0.37]	< .001	
Morality					0.04	[0.00, 0.07]	.029	
Happiness					0.16	[0.12, 0.20]	< .001	
Goal attainment					0.25	[0.20, 0.29]	< .001	
Step 4								.49
Low level	0.39	[0.33, 0.46]	< .001	.018	0.02	[–0.06, 0.11]	.586	
Suboptimal level	0.51	[0.46, 0.57]	< .001	.007	0.31	[0.23, 0.38]	< .001	
Morality	–0.04	[–0.11, 0.03]	.29	.076	0.04	[0.00, 0.08]	.049	
Happiness	0.41	[0.35, 0.48]	< .001	.054	0.16	[0.12, 0.21]	< .001	
Goal attainment	0.45	[0.39, 0.51]	< .001	.048	0.23	[0.18, 0.28]	< .001	
Difficulty	0.34	[0.28, 0.39]	< .001	.006	0.12	[0.08, 0.17]	< .001	
Controllability	–0.18	[–0.24, –0.12]	< .001	.012	0.05	[0.01, 0.10]	.014	

Note. τ^2 = slope variance.

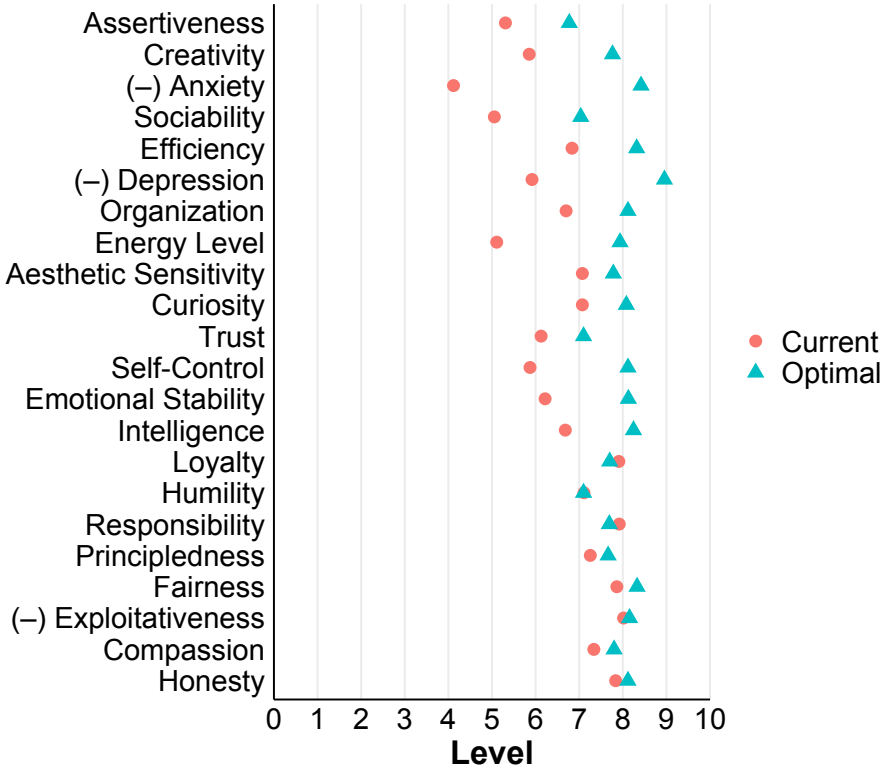
Suboptimal Levels. Next, we tested whether participants in the Supplemental Study perceive smaller discrepancies between their optimal and current levels for more morally relevant traits. As shown in Figure S8, this was the case, β = –0.06, 95% CI [–0.12, –0.01], *p* = .017. A non-preregistered analysis also showed that people perceive having higher levels of more morally relevant traits, β = –0.12, 95% CI [–0.19, –0.06], *p* < .001.

In single-predictor models predicting change goals (see Table S14), participants in the Supplemental Study were more interested in improving the traits that they perceived deficiencies on, either in terms of low levels or suboptimal levels. The two strongest predictors were beliefs about consequences for goal attainment and happiness. Surprisingly, however, in contrast to

Studies 1 and 2a, Supplemental Study participants were not detectably less interested in improving traits that they personally considered to be more morally relevant. In line with prior work (Sun & Goodwin, 2020), however, a non-preregistered by-traits analysis clearly showed that Supplemental Study participants were generally less interested in improving the traits that were generally considered to be more morally relevant, $r = -.73$, $p < .001$.

Figure S8

Discrepancy Between Trait and Optimal Levels in the Supplemental Study



Note. Traits are ordered from least morally relevant (top) to most morally relevant (bottom).

When all predictors were entered into the model, Supplemental Study participants were more interested in improving traits that they perceived having suboptimal levels of and that they believed would help them attain their goals and increase their happiness to a greater extent (recall that we did not measure beliefs about consequences for meaning, connected and social

status in this study). Participants were also somewhat more interested in improving traits that they believed would be more difficult and more within their control to change. Interestingly, in this sample, after controlling for beliefs about happiness and goal attainment consequences, participants were slightly more interested in improving more morally relevant traits. In short, these results continue to show that goal attainment and happiness are important predictors of change goals, even after accounting for various other explanations.

Discussion

Broadly speaking, the results continued to show that participants in Study 2a and the Supplemental Study perceive a tradeoff between the extent to which various trait improvements would improve morality vs. personal fulfillment, even when the definition of morality no longer mentioned self-sacrifice. Beliefs about consequences for goal attainment and happiness also continued to predict change goals after controlling for difficulty, controllability, and discrepancy between current and optimal levels of traits.

However, there were a few findings that differed from Study 2a. When using person-specific estimates from participants in the Supplemental Study, we did not find a detectable tradeoff between morality and happiness. We also did not find that participants in the Supplemental Study were less interested in improving the traits that they personally considered to be more morally relevant.

Given these ambiguities, we decided to conduct Study 2b, which was a very close replication of Study 2a, except that we removed the potential "self-sacrifice" confound and kept both operationalizations of perceived deficiencies (as in the Supplemental Study). Given that Study 2b had three times the number of participants as the Supplemental Study and was

conducted in-person, we consider this to be a more definitive test of whether people truly perceive a tradeoff between well-being and morality in the context of self-improvement.

References

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*(3), 497–529. https://doi.org/10.1037/0033-2909.117.3.497

Baumeister, R. F., Vohs, K. D., Aaker, J. L., & Garbinsky, E. N. (2013). Some key differences between a happy life and a meaningful life. *The Journal of Positive Psychology*, *8*(6), 505–516. https://doi.org/10.1080/17439760.2013.830764

Butler, J., & Kern, M. L. (2016). The PERMA-profiler: A brief multidimensional measure of flourishing. *International Journal of Wellbeing*, *6*(3), Article 3. https://doi.org/10.5502/ijw.v6i3.526

Buuren, S. van, & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3). http://www.jstatsoft.org/v45/i03/paper

Deci, E. L., & Ryan, R. M. (2012). Motivation, personality, and development within embedded social contexts: An overview of self-determination theory. In R. M. Ryan (Ed.), *The Oxford handbook of human motivation* (pp. 85–107). Oxford University Press. https://psycnet.apa.org/record/2012-03687-006

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology*, *93*(5), 880–896. https://doi.org/10.1037/0022-3514.93.5.880

DeYoung, C. G., & Tiberius, V. (2022). Value fulfillment from a cybernetic perspective: A new psychological theory of well-being. *Personality and Social Psychology Review*, 3–27. https://doi.org/10.1177/10888683221083777

Dwyer, R. J., Dunn, E. W., & Hershfield, H. E. (2017). Cousins or conjoined twins: How different are meaning and happiness in everyday life? *Comprehensive Results in Social Psychology*, *2*(2–3), 199–215. https://doi.org/10.1080/23743603.2017.1376580

Furr, R. M. (2008). A framework for profile similarity: Integrating similarity, normativeness, and distinctiveness. *Journal of Personality*, *76*(5), 1267–1316. https://doi.org/10.1111/j.1467-6494.2008.00521.x

Furr, R. M., Prentice, M., Hawkins Parham, A., & Jayawickreme, E. (2022). Development and validation of the Moral Character Questionnaire. *Journal of Research in Personality*, *98*, 104228. https://doi.org/10.1016/j.jrp.2022.104228

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

Gray, K., DiMaggio, N., Schein, C., & Kachanoff, F. (2022). The problem of purity in moral psychology. *Personality and Social Psychology Review*. https://doi.org/10.1177/10888683221124741

Moore, A. K., Munguia Gomez, D. M., & Levine, E. E. (2019). Everyday dilemmas: New directions on the judgment and resolution of benevolence–integrity dilemmas. *Social and Personality Psychology Compass*, *13*(7), e12472. https://doi.org/10.1111/spc3.12472

Oishi, S., & Westgate, E. C. (2022). A psychologically rich life: Beyond happiness and meaning. *Psychological Review*, *129*, 790–811. https://doi.org/10.1037/rev0000317

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Wiley.

Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*, *57*, 1069–1081. https://doi.org/10.1037/0022-3514.57.6.1069

Sherman, R. A., & Serfass, D. G. (2015). The comprehensive approach to analyzing multivariate constructs. *Journal of Research in Personality*, *54*, 40–50. https://doi.org/10.1016/j.jrp.2014.05.002

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, *113*, 117–143. https://doi.org/10.1037/pspp0000096

Sun, J., & Goodwin, G. P. (2020). Do people want to be more moral? *Psychological Science*, *31*(3), 243–257. https://doi.org/10.1177/0956797619893078

Walker, L. J., & Frimer, J. A. (2007). Moral personality of brave and caring exemplars. *Journal of Personality and Social Psychology*, *93*(5), 845–860. https://doi.org/10.1037/0022-3514.93.5.845