Received: 13 October 2022 | Revised: 12 January 2023 | Accepted: 14 January 2023

DOI: 10.1111/jopy.12812

ORIGINAL ARTICLE

WILEY

How and why people want to be more moral

Jessie Sun[1] | **Joshua Wilt**[2] | **Peter Meindl**[3] | **Hanne M. Watkins**[4] | **Geoffrey P. Goodwin**[4]

[1]Department of Psychological and Brain Sciences, Washington University in St. Louis, St. Louis, Missouri, USA

[2]Department of Psychological Sciences, Case Western Reserve University, Cleveland, Ohio, USA

[3]United States Military Academy, West Point, Utah, USA

[4]Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania, USA

Correspondence
Jessie Sun, Department of Psychological and Brain Sciences, Washington University in St. Louis, 1 Brookings Drive, St. Louis, MO 63130, USA.
Email: jessie.sun@wustl.edu

Abstract

Objective: What types of moral improvements do people wish to make? Do they hope to become more good, or less bad? Do they wish to be more caring? More honest? More loyal? And why exactly do they want to become more moral? Presumably, most people want to improve their morality because this would benefit others, but is this in fact their primary motivation? Here, we begin to investigate these questions.

Method: Across two large, preregistered studies (N = 1818), participants provided open-ended descriptions of one change they could make in order to become more moral; they then reported their beliefs about and motives for this change.

Results: In both studies, people most frequently expressed desires to improve their compassion and more often framed their moral improvement goals in terms of amplifying good behaviors than curbing bad ones. The strongest predictor of moral motivation was the extent to which people believed that making the change would have positive consequences for their own well-being.

Conclusions: Together, these studies provide rich descriptive insights into how ordinary people want to be more moral, and show that they are particularly motivated to do so for their own sake.

KEYWORDS

moral character, moral improvement, moral motivation, personality change, well-being

1 | INTRODUCTION

Religion and philosophy have long provided prescriptions for what it means and what it takes to be an ethical, virtuous person (Grudem, 2018; Telushkin, 2011; Wilburn, 2010). The objective of moral improvement has also made its way into contemporary educational, business, and bioethical settings. For example, moral education and character education programs aim to cultivate moral reasoning skills and virtuous traits in students (Althof & Berkowitz, 2006; Lavy, 2020). Business ethics classes aim to develop more ethical future business leaders. More recently, there has been increasing urgency around the topic of moral enhancement by biomedical means to improve traits that would help preserve humanity's long-term future (Persson & Savulescu, 2012). Yet, despite these various societal efforts to encourage moral cultivation, little is known about how ordinary people think about moral improvement. Here, we begin to address this gap by providing a rich, descriptive investigation of how and why people want to be more moral.

Past research shows that people generally want to improve their personality traits (Baranski et al., 2021; Hudson & Fraley, 2016). When asked to prioritize, however, people are much more interested in improving traits

The views expressed herein are those of the authors and do not reflect the position of the United States Military Academy, the Department of the Army, or the Department of Defense.

that are typically thought of as nonmoral (e.g., anxiety, sociability, productiveness), compared with prototypically moral traits (e.g., honesty, compassion, fairness; Sun & Goodwin, 2020). In other words, most people do not seem to be particularly interested in being more moral.

However, previous studies show that conclusions about which traits people want to change can depend on whether people report their change goals using predefined personality inventories or in an open-ended way (Baranski et al., 2021; Miller, 2022). In Sun and Goodwin's (2020) study, moral change goals were measured by asking people to report the extent to which they wished to increase, decrease, or stay the same on the broad moral character traits of general morality, respectfulness, fairness, honesty, loyalty, compassion, responsibility, and purity. Goals to change each trait were measured with two items per trait, such as "I want to be helpful and unselfish with others," "I want to be a fair person," and "I want to be honest even when it's hard." However, by relying on a relatively limited conception of moral improvement, Sun and Goodwin's study arguably does not reveal very much about how people naturally conceptualize moral improvement. Moreover, to the extent that people think about moral improvement in a different way than was measured by Sun and Goodwin, that study might have underestimated people's moral motivation.

For example, perhaps people wish to improve different moral traits (e.g., gratitude, bravery, humility) than the ones that were included in the previous study. Moreover, broad, decontextualized moral character traits are only one level at which a person's moral personality may be described (McAdams, 1995). People might also have more contextualized moral strivings, in the form of goals to improve specific moral behaviors (e.g., eating less meat) or to improve moral conduct in specific contexts (e.g., in relation to specific moral causes) or within specific roles (e.g., being more patient with one's children).

In addition, all but one of the items used to measure moral change goals in Sun and Goodwin's (2020) study were framed in terms of increasing positive moral tendencies. Yet, grounded in the well-established distinction between approach and avoidance motivation (Elliot & Thrash, 2002), there is evidence for two systems of moral motivation: prescriptive morality focuses on what we *should* do, involves activating "good" behaviors to attain positive outcomes, and is seen as being more discretionary and creditworthy, whereas proscriptive morality focuses on what we should *not* do, involves inhibiting immoral actions to avoid negative actions, and is seen as being more obligatory and blameworthy (Janoff-Bulman et al., 2009). This implies that there are two ways to conceptualize moral improvement: in terms of amplifying (i.e., starting or increasing) positive tendencies (e.g., donating more to charity), or in terms of curbing (i.e., stopping or decreasing) negative tendencies (e.g., not gossiping). But, we do not currently know whether people more typically characterize their desired moral improvements in prescriptive or proscriptive terms.

Finally, although Sun and Goodwin (2020) found that people rarely *prioritized* moral improvements, most people nevertheless indicated a preference for improving traits such as honesty, compassion, and fairness (as opposed to staying the same or decreasing on these traits). Moreover, 5%–10% of people did prioritize goals to become more moral (Sun & Goodwin, 2020). This shows that most people think that being more moral would be desirable (even if it was not a top priority), and some people even prioritize it. Accordingly, it seems worthwhile to better understand what motivates or discourages people from becoming more moral, and which kinds of moral improvements are particularly appealing. For example, for whose sake do people most want to change? Do people think that moral improvements would be particularly difficult, or even within their control? And, are people more motivated to improve when they perceive greater benefits to others or to themselves?

Understanding how people envision their own moral improvement has both theoretical and practical value. First, the details of people's moral aspirations can reveal which domains of morality are particularly relevant in their daily lives and which kinds of moral shortcomings are particularly common (from their perspective). This would provide important descriptive information on the scope and limits of typical moral motivation, which can be used to inform theories of moral functioning. Such information could also have implications for debates about the ethics of moral bio-enhancement. For example, if the moral traits that philosophers believe should be enhanced are the same traits that ordinary people frequently want to improve, there would be fewer ethical concerns related to freedom, identity, and autonomy (and vice versa; Specker et al., 2014). Finally, such knowledge might inform our understanding of which factors might be most important to intervene on to inspire moral change. For instance, if people are more motivated to be more morally good than to be less morally bad, this might suggest that moral change messages that are framed in approach terms might be more effective than messages that are framed in avoidance terms. Similarly, if people believe that moral improvements are relatively out of their control or relatively difficult, then increasing perceptions of controllability or decreasing perceptions of difficulty might be a promising first step.

1.1 | The present research

In sum, our aims are to investigate three aspects of what moral improvement looks like in the minds of ordinary people. First, we describe the framing, specificity, and trait content of people's moral improvement goals.

Second, we explore people's beliefs about the difficulty, controllability, and well-being consequences of moral improvement. Third, we aim to understand what motivates people to be more moral. To do so, we conducted two large, preregistered studies. Study 1 provides an initial investigation among a sample of visitors of YourMorals.org. Study 2 addresses methodological limitations of Study 1 and provides a conceptual replication among a sample of CloudResearch participants.

Data collection procedures for Study 1 were approved by the Institutional Review Board (IRB) at the University of Southern California (IRB ID: UP-07-00393; Study Title: Morality Studies). Data collection procedures for Study 2 were approved by the IRB at the University of Pennsylvania (IRB ID: 831767; Study Title: Moral Change Goals). Preregistrations for both studies are available at https://osf.io/g3qeb and https://osf.io/d9fqa, respectively. All materials (including measures outside the scope of the current project) and the data and code required to reproduce the results reported in this manuscript are available at https://osf.io/qzs8g/.

2 | STUDY 1

2.1 | Method

2.1.1 | Participants

Participants were visitors to YourMorals.org, an online platform that was developed to collect data for moral psychology studies. At any given time, dozens of studies were active. Visitors to the website were able to see a list of all active studies and could choose to participate in as many studies as they wanted to. A total of 1557 responses were collected between May 26, 2018 and September 13, 2018. However, because we did not start this project with firm hypotheses, we split the dataset into exploratory ($n = 487$) and confirmatory subsets ($n = 1070$). We used the exploratory subset to develop and refine the scope of the project and the coding scheme described below. We then preregistered our final coding scheme and analysis plan for the confirmatory subset. We report the results of the confirmatory subset in the main text.

For the confirmatory subset, we excluded 116 participants who did not provide a valid response to the key open-ended question about their desired moral improvement (described below). As preregistered, invalid responses included (a) nonresponses, (b) nonsensical responses, (c) responses that indicated that the participant did not want to be more moral, and (d) responses that were not in English. After applying these preregistered exclusions, the final subset involved 954 participants (339 women, 612 men, 3 not reported) who reported being 36.83 years old on average (SD = 16.35). The overwhelming majority of participants were based in the U.S. ($n = 691$; 72.43%), followed by the U.K. ($n = 65$; 6.81%), Canada ($n = 56$; 5.87%), and Australia ($n = 35$; 3.67%). The remaining participants were based in 45 other countries (9.75%) or did not report their country (1.47%).

2.1.2 | Measures

Goal content variables

Participants were asked to "please spend some time describing one and only one thing you want to change about yourself in order to become more morally good," using at least 10 words. Expert coders and research assistants coded the responses for the variables described below. Instructions for the coding tasks (including detailed definitions and distinctions between different trait categories) are available at https://osf.io/qzs8g/.

Breadth of change goals. Two research assistants coded whether the response reflected a goal to change a specific or contextualized behavior (e.g., eating meat, being kinder to specific others) or a broad trait (e.g., being more compassionate). Interrater agreement was $\kappa = .53$. A third research assistant adjudicated between coding disagreements (14.88%).

Amplifying versus curbing. Two research assistants coded whether they would characterize the change the participant wants to make as stopping (or decreasing) an undesirable tendency, starting (or increasing) a desirable tendency, both, or truly ambiguous. Interrater agreement was $\kappa = .83$. A third research assistant adjudicated between coding disagreements (8.91%). Truly ambiguous responses were treated as missing values.

Trait content categories. We aimed to develop a coding scheme that captured a comprehensive set of moral and nonmoral traits. To do so, we initially included 27 trait categories from several taxonomies of traits and moral values (for details, see Supplemental Material). We then used the results from the exploratory subset to simplify the coding scheme by dropping categories with low base rates and combining conceptually similar categories (for details, see the Supplemental Material). The final 20 traits and the taxonomies from which they were derived are reported in Table 1. Expert coders (the first and second authors) coded the responses into either the socially desirable (e.g., more compassionate, less reactive) or socially undesirable (e.g., less compassionate, more reactive) pole of one of these 20 traits, plus an "Other" category for any responses that could not be captured by this coding scheme (for example responses, see Table 2). This yielded a total of 41 mutually

TABLE 1 Definitions and sources for the trait content coding categories.

Trait	Definition	Source			
		BFI-2	VIA	MFT	HEXACO-H
Agreeableness					
Compassion	Care/concern for others' well-being, including behaviors such as helpfulness and generosity	Compassion	Love, Kindness	Harm/Care	
Respectfulness	Treating others with regard for their personal preferences and rights, while inhibiting antagonistic and aggressive impulses	Respectfulness			
Trust	Positive generalized beliefs about others	Trust			
Gratitude	Feeling and expressing a deep sense of thankfulness in life and for others		Gratitude		
Conscientiousness					
Responsibility	Commitment to meeting duties and obligations	Responsibility			
Productiveness	Work ethic and persistence while pursuing goals	Productiveness	Perseverance		
Organization	Preference for order and structure	Organization			
Self-control	Overriding impulses/resisting temptations; exercising self-discipline		Self-regulation		
Honesty–Humility					
Honesty	Avoiding fraud and corruption; being genuine with others		Honesty		Fairness, Sincerity
Humility	Being modest and unassuming; being uninterested in possessing lavish wealth and signs of high social status		Humility		Modesty, Greed Avoidance
Purity/Spirituality	Sacredness and spiritual purity; connecting with the transcendent		Spirituality	Purity	
Open-mindedness	Analytically evaluating ideas; seeing the bigger picture and multiple sides of an issue; intellectual humility; intellectual interests and enjoyment of thinking; creativity and originality	Intellectual Curiosity, Creative Imagination	Curiosity, Love of Learning, Perspective, Creativity		
Extraversion					
Sociability	Desire to socially approach and engage with others	Sociability			
Energy Level	Positive affect and activity level	Energy Level	Zest		
Bravery	Acting on convictions and facing fears; includes physical, psychological, and moral bravery; willingness to express personal opinions and goals in social situations	Assertiveness	Bravery		
Emotional stability					
Depression	Tendencies toward depression, sadness, insecurity, and pessimism	Depression			
Anxiety	The tendency to experience anxiety, fear, and stress	Anxiety			
Reactivity	The tendency to experience volatile mood swings; being easily angered or irritated (as opposed to being calm and patient in the face of frustration or adversity)	Emotional Volatility	Self-regulation		
Other					
Fairness	Treating people justly and without prejudice		Fairness	Fairness	
Loyalty	Patriotism, self-sacrifice for the group, vigilance for traitors		Teamwork	Loyalty	

Note: BFI-2, Big Five Inventory-2 (Soto & John, 2017); VIA, Values in Action Survey of Character Strengths (Peterson & Seligman, 2004); MFT, Moral Foundations Theory (Graham et al., 2013); HEXACO-H, Honesty–Humility domain of the HEXACO trait taxonomy (Lee & Ashton, 2018).

TABLE 2 Examples of moral improvement goals within each trait category.

Trait	Study 1 (expert codings)	Study 2 (self-codings)
Agreeableness		
Compassion	To practice more kindness towards others in daily life.	I want to donate more of my time and money to charity.
Respectfulness	I want to be respectful of others' beliefs even when they challenge/condemn me.	I want to change the way I look at women. I want to have an attitude of respect for their potential as a person instead of basing my opinion of them on their looks.
Trust	I would like to give people the benefit of the doubt more often. I'm very likely to write people off quickly if I think they are dumb, annoying, etc.	I want to learn to be more forgiving to those who have wronged me.
Gratitude	Be patient and grateful for the blessings I have and be good to others.	Be more appreciative of family.
Conscientiousness		
Responsibility	I want to be more responsible about my actions. They tend to be selfish and reckless.	I want to follow through with my commitments to other people more.
Productiveness	Get a job to be productive and contribute to society.	I want to be a hard, diligent worker, and make a positive contribution to society.
Organization	I want to budget my time better such as less time on frivolous things and more on serious helpful issues.	I want to be more consistent in my workouts and cleaning habits.
Self-control	I would increase my willpower so that I would be more capable of acting out what is right.	I want to drink less alcohol.
Honesty–Humility		
Honesty	I would like to be more honest with myself and with others.	I want to lie less. I feel like I tell white lies because it's easy, and I want to get out of that habit.
Humility	I want to stop feeling self-important and looking down on other people.	I want to stop thinking I am better than other people.
Purity/Spirituality	To deepen my relationship with Jesus Christ through study, prayer, and spreading the gospel.	I want to stop looking at pornography.
Open-mindedness	Become more well-read about the ideas of those who do not agree with my philosophies or my particular political views.	Be more socially educated to have a better ability to consider many different views, giving me a better chance to make morally correct decisions.
Extraversion		
Sociability	I think introversion gets in the way of me being more involved with others, to the point where I am not as involved in my church or with my family and friends as I could.	I would like to help the elderly members of my local community, I think a great way to do that would be to regularly visit an elderly person that is lonely or offer help such as picking up groceries or taking the trash out.
Energy Level	I wish I felt more energetic about physically working to improve the situation for myself and others around me. I feel tired and worn down all the time, and wish I were more energetic and motivated.	The one thing that I want to change about myself is the fact that I want to not be such a glutton. I love to eat and eat a lot.
Bravery	I wish I was braver in standing up for what I know is right and not care so much about being criticized for it.	I wish I had more moral courage so that I could speak out when I see an injustice occur.
Emotional stability		
Depression	I want to change my pessimistic outlook on the world and life.	I would like to change my pessimistic viewpoint of life.
Anxiety	I want to be able to make better choices under upsetting or stressful situations.	I think my anxiety has stunted my growth and I could become a morally better person if I got the help that I need.
Reactivity	I want to be less volatile, angry, and impatient...	I want to better control my temper.

TABLE 2 (Continued)

Trait	Study 1 (expert codings)	Study 2 (self-codings)
Other		
Fairness	I would like to change my thoughts and opinions about race. I was raised in a primarily white town and as a result, I notice that despite my best efforts to educate myself on racial issues and struggles, I still find myself holding pre-assumed racial biases...	I want to give my part of "what's fair," I feel kind of hypocritical when my wife asks if I can bring her coffee, but I tell her we cannot really afford that, but end up getting a coffee myself for example.
Loyalty	Showing more loyalty to people and not discussing them when not present to defend themselves.	I wish I could stop cheating on my boyfriend.
Other	Do more of what makes me happy.	I wish I could go back in time and never get addicted to pills as I did some things that I wasn't proud of.

Note: Ellipses indicate that the examples have been truncated. Examples have been edited for grammatical and spelling errors. This table only includes changes in the socially desirable direction, as changes in the socially undesirable direction (e.g., goals to become less compassionate) were extremely rare.

exclusive categories. Interrater agreement was $\kappa = .63$. A third expert coder (the fourth author) adjudicated between coding disagreements (33.33%).

These trait categories are, in principle, orthogonal to the amplifying versus curbing variable. For example, goals to become more compassionate (for example) could be framed in terms of increasing positive tendencies (e.g., "To spend more time serving my community") or decreasing negative tendencies (e.g., "Stop eating meat"); similarly, goals to become less reactive (for example) could be framed in terms of increasing positive tendencies (e.g., "I would like to become more patient and tolerant of other people's views and opinions") or decreasing negative tendencies (e.g., "I want to stop taking my anger out on others"). In addition, although one of our goals was to examine whether people think about moral improvement in terms of changing broad traits or in terms of a narrower context (e.g., specific behavior or relationship), almost every thought, feeling, or behavior falls somewhere within our trait content taxonomy. Therefore, it is possible to assess the trait content (e.g., more honest) of both a broad trait change (e.g., "Learn to be honest and tell the truth in as many circumstances as possible") and a contextualized change (e.g., "I would like to be more honest with myself").

We report the prevalence of moral improvement goals that fell within each of the narrow trait categories (plus the "Other" category), to provide a comprehensive description (see Table S2 and Figure 1). However, for the regression analyses, as preregistered, we reduced the 41 categories into the most parsimonious set of predictors based on conceptual and empirical grounds (see Table 1). First, because almost all of the changes that reflected one of the 20 traits were in a socially desirable direction, the rare exceptions (0.10% goals to become less trusting and 0.31% goals to become less compassionate) were grouped within the Other category. For the socially desirable changes, based on expert knowledge of personality taxonomies, we judged that almost all of the narrower traits could be grouped within one of the Big Five domains or Honesty–Humility. The exceptions were Purity/Spirituality, Loyalty, and Fairness. Because Purity/Spirituality was mentioned reasonably frequently in the exploratory data (see Supplemental Material), we kept it as its own category. Loyalty and Fairness were each mentioned less than 1.2% of the time in the exploratory data. Thus, for parsimony, we grouped these under the Other category (along with any trait that did not reflect one of the 20 traits in the coding scheme, as well as any goals to change in the socially undesirable direction). Interrater agreement for the trait domains was $\kappa = .68$.

Goal perception variables

Motivation. Participants reported the extent (if at all) to which they are motivated to make this change (1 = *Not at all motivated*, 6 = *Extremely motivated*).

Beneficiaries. In response to the question, "For whose sake are you MOST motivated to make this change (i.e., whose interests do you most have at heart in wanting to make this change)?" participants selected one entity out of the following list (adapted from Crimston et al., 2016): Myself, Immediate family members, Members of my extended family, My closest friends, Distant friends, My acquaintances, People I have met but don't know well, My country/the people in my country, My continent/the people in my continent, The people on all continents, Mammals on all continents, Many types of animals on earth (fish, reptiles, mammals, etc.), Many types of living things in the universe (trees, animals, etc.), All natural things in the universe including rocks, and All things in existence.

Costs and benefits for the self and others. Participants selected the person for whom they expected this change would have the most positive consequences out of three

mutually exclusive options: Me, Other person/people, Something else (e.g., animals, the environment, etc.).

Participants also reported their beliefs about the extent to which making the change would have positive (1 = *Not at all positive*, 6 = *Extremely positive*) and negative (1 = *Not at all negative*, 6 = *Extremely negative*) consequences for themselves. Using two separate items, they also reported their beliefs about the extent to which making the change would have positive and negative consequences for "another person, other people, and/or something else (e.g., animals, the environment)," using the same response scales as above. For the analyses reported in Table 4, we combined the positive and negative consequences variables (for the self and for others, respectively) into two overall consequences variables (for the self and for others) by averaging the positive and negative items together (after reverse-scoring negative consequences).

Controllability. Participants reported the extent to which they thought making this change was within their control (1 = *Not at all within my control*, 6 = *Extremely within my control*).

Difficulty. Participants reported how difficult it would be for them to make the change (1 = *Not at all difficult*, 6 = *Extremely difficult*).

2.1.3 | Data analyses

Analyses were conducted in R (R Core Team, 2022). Not all participants completed all self-report questions listed above. As preregistered, we used all available data. Thus, sample sizes ranged between 842 and 954 for a given analysis. For regression analyses, the broad versus specific variable (reference group = specific) and the amplifying versus curbing variable (reference group = curbing) were both dummy-coded. That is, the regression coefficients represent the difference between the mean of the group coded 1 (e.g., goals to amplify positive tendencies) and the reference group coded 0 (e.g., goals to curb negative tendencies). For the trait domain predictors, there was no obvious reference group. To better facilitate comparisons of the traits with one another, we therefore used effect coding (contrasting category = Other). For each effect variable (e.g., Agreeableness), the focal category (for this example: Agreeableness) was coded as 1, the contrasting category (always Other) was coded as −1, and all other categories (for this example: Conscientiousness, Honesty–Humility, Purity/Spirituality, Open-Mindedness, Extraversion, and Emotional Stability) were coded as 0. The resulting regression coefficients for the effect-coded variables therefore represent the deviation from the unweighted grand mean (i.e., the mean of the means of each of the eight trait categories) of the dependent variable. In this way, the effect of a particular trait category (e.g., Agreeableness) can be interpreted as the difference between that category and the "average" moral goal (as if the number of observations within each trait category was balanced).

2.2 | Results

2.2.1 | How do people conceptualize moral improvement?

First, we aimed to understand how people typically conceptualize moral improvement. Do people typically think of moral improvement in terms of changing specific behaviors or broad character traits, and in terms of curbing negative qualities or amplifying positive ones? Which moral traits are people most frequently interested in improving in themselves?

Codings of the open-ended responses showed that people overwhelmingly reported desires to improve broad character traits (81.24%; e.g., "be more honest with myself and others," "be more selfless," "genuinely care about other people and their well-being more"), rather than specific or contextualized behaviors (18.76%; e.g., "Gossip less," "Give more to charity," "To be less selfish with my partner"). People largely reported desires to amplify positive tendencies (66.35%; e.g., "spend more time serving my community"), rather than to curb negative ones (16.88%; e.g., "Be less judgmental") or to both curb negative tendencies and amplify positive ones (16.56%; e.g., "Increase patience and decrease reactivity"). The remaining responses were ambiguous (0.21%).

As shown in Figure 1, people reported a diverse range of moral improvement goals that spanned all 20 trait categories as well as traits that were not included in the coding scheme (and were therefore coded into the "Other" category). These goals almost exclusively reflected the socially desirable poles of each trait (i.e., low levels of anxiety, depression, and reactivity, and high levels of all other traits), with the exception of a small proportion of goals to be less trusting (0.10%) and less compassionate (0.31%). Goals to become more compassionate represented almost one-fifth of the moral improvement goals (19.08%), followed by goals to become more open-minded (13.31%), more productive (9.43%), less reactive (8.81%), more honest (7.97%), and more brave (6.71%). Interestingly, several goals that are less directly or uniquely relevant to morality (becoming less depressed, more sociable, or more productive) were mentioned more often than goals that are more directly relevant to morality (becoming more humble, fair, loyalty, or grateful).

2.2.2 | What do people believe about the process and consequences of moral improvement?

Next, we explored people's self-reported beliefs about the processes and consequences of moral improvement. As shown in Table 3, although people thought that making their desired moral change would be somewhat difficult, they also thought it was very much within their control. Intriguingly, when asked for whom they expected this change would have the most positive consequences, over half of participants selected "Me" (52.75%), followed by "Other person/people" (42.09%), and "Something else" (5.16%). However, nonpreregistered analyses suggested that these perceptions varied depending on the trait in question (see Figure S2); for example, people thought that becoming more compassionate would primarily benefit others (66.46%), whereas becoming more honest would primarily benefit themselves (69.57%). People also believed that moral improvements would have very positive consequences and only little negative consequences for themselves (see Table 3). These findings suggest that people overwhelmingly forecast substantial benefits

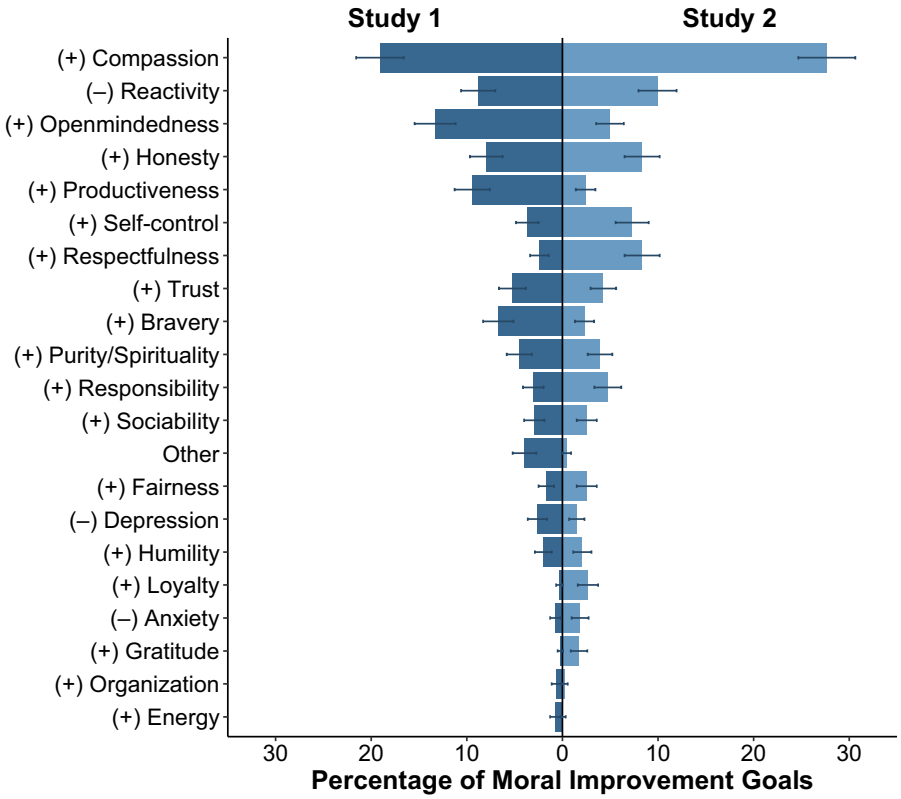


FIGURE 1 Trait categorizations of the most frequently mentioned moral improvement goals in Studies 1 and 2. Goals to change in a socially undesirable direction (0.41%) were recoded as "Other" in Study 1. Goals are ordered from most to least frequent, averaged across the two samples (weighted equally). Error bars denote 95% confidence intervals.

TABLE 3 Descriptive statistics and correlations among continuous variables (Study 1).

Variable	*M*	*SD*	1.	2.	3.	4.	5.	6.
1. Change motivation	4.27	1.11						
2. Positive consequences for self	4.81	1.21	.42					
3. Negative consequences for self	1.85	1.15	−.15	−.34				
4. Positive consequences for others	4.61	1.22	.26	.27	−.05			
5. Negative consequences for others	1.39	0.87	−.03	−.08	.40	−.11		
6. Difficulty	4.36	1.13	.00	.06	.21	.02	.18	
7. Controllability	4.82	1.11	.19	.12	−.10	.08	−.04	−.30

Note: $rs \geq |.08|$ were detectable at $p < .05$. All variables were measured on a 1–6 scale.

and few downsides of becoming more moral, for both themselves and for others.

Next, we explored whether people believed that some changes would be more or less difficult and within their control than others, and whether some changes would result in more or less favorable overall consequences for themselves and for others. As shown in Table 4, people believe that it is less difficult to amplify positive tendencies than to curb negative tendencies. Those who reported agreeableness- or openness-related changes also believed that such changes would be less difficult to make, whereas those who reported extraversion-related changes believed that such changes would be more difficult to make (compared with the unweighted grand mean across all trait domain categories). We did not find any detectable differences in the perceived controllability of different kinds of moral changes.

As shown in Table 4, people believed that broad trait changes (compared with specific or contextualized changes) would result in more favorable overall consequences for themselves. They also believed that improving their purity would result in more favorable consequences, whereas improving their honesty–humility and their extraversion would result in less favorable consequences for both themselves and for others (compared with the unweighted grand mean). In addition, people believed that improving their conscientiousness would result in more favorable consequences for themselves (compared with the unweighted grand mean). Finally, people believed that agreeableness-related changes would result in less favorable consequences for themselves, whereas such changes would result in more favorable consequences for others (compared with the unweighted grand mean).

2.2.3 | What motivates people to be more moral?

Finally, we aimed to understand what motivates people to be more moral. Presumably, people would want to improve their morality to benefit others. However, when asked to indicate for whose sake they are most motivated to make the moral change they mentioned, almost half of the participants reported that they were most motivated to change for their own sake (48.30%). This was followed by immediate family members (16.02%), people on all continents (7.49%), all things in existence (6.55%), their country or the people in their country (5.61%), their closest friends (5.38%), acquaintances (2.46%), and people they have met but don't know well (2.22%). All other categories were selected less than 2% of the time.

Next, we predicted motivation to enact the desired change from the content of the goals, as well as self-reported beliefs about the consequences, difficulty, and controllability of the change. We did so separately for the set of coded variables ($R^2 = .04$), separately for the set of self-reported variables ($R^2 = .23$), and finally, with all variables included in the model ($R^2 = .25$). As shown in Table 5, the final model showed that people who reported agreeableness- and conscientiousness-related improvements were less motivated to make those improvements (compared with the unweighted grand mean). People who said that they were most motivated to change for their own sake were slightly less motivated than people who said that they were most motivated to change for others' sake. People who believed that the change would have greater positive consequences for both the self and for others, and who believed that making the change was more within their control, were more motivated to make those improvements (compared to the unweighted grand mean).

Because beliefs about the extent to which the change would result in positive consequences for the self appeared to be a stronger predictor of motivation to make the desired change than were beliefs about the extent to which the change would result in positive consequences for others, we conducted a nonpreregistered analysis to explore whether these two regression coefficients were significantly different. To do so, we ran another regression analysis with the same set of predictors, but constrained the regression coefficients for positive consequences for the self and for others to equality. This model fit significantly worse than the model in which the coefficients were free to vary, $\Delta\chi^2\,(df = 1) = 13.82$, $p < .001$. Thus, anticipated positive consequences for oneself more strongly predict moral motivation than do anticipated positive consequences for others.

2.3 | Discussion

In sum, in Study 1, we found that people primarily think of moral improvement in terms of improving broad moral character traits (rather than specific or contextualized behaviors) and in terms of amplifying positive qualities (rather than curbing negative ones). The most common moral improvement goal was to become more compassionate, followed by the goal to become more open-minded. Finally, surprisingly, we found that people view themselves as the primary beneficiaries of their own moral improvements, and that they are relatively more motivated by the potential personal benefits of a given moral improvement than by the potential benefits to others.

However, there were several aspects of the protocol and the sample that could limit the generalizability of these findings. The wording of the instructions could

TABLE 4 Multiple regression analyses predicting perceived difficulty, controllability, and overall consequences from goal content (Studies 1 and 2).

	Difficulty			Controllability			Consequences (self)			Consequences (others)		
	β	95% CI	*p*	*β*	95% CI	*p*	*β*	95% CI	*p*	*β*	95% CI	*p*
Study 1												
Broad vs. specific	0.13	[−0.05, 0.30]	.150	0.01	[−0.16, 0.19]	.896	**0.40**	**[0.23, 0.57]**	**<.001**	0.04	[−0.14, 0.21]	.673
Amplifying vs. curbing												
Amplifying	**−0.32**	**[−0.50, −0.13]**	**.001**	0.06	[−0.14, 0.25]	.575	−0.07	[−0.26, 0.11]	.451	0.00	[−0.19, 0.19]	.970
Both	−0.12	[−0.35, 0.11]	.307	−0.07	[−0.31, 0.16]	.547	−0.07	[−0.30, 0.16]	.561	0.10	[−0.13, 0.34]	.401
Trait content												
Agreeableness	**−0.32**	**[−0.45, −0.19]**	**<.001**	−0.09	[−0.23, 0.04]	.173	**−0.18**	**[−0.31, −0.05]**	**.008**	**0.14**	**[0.01, 0.28]**	**.038**
Conscientiousness	0.12	[−0.03, 0.28]	.120	0.04	[−0.12, 0.21]	.598	**0.17**	**[0.01, 0.32]**	**.037**	−0.01	[−0.17, 0.15]	.925
Honesty–Humility	0.14	[−0.06, 0.33]	.171	0.17	[−0.03, 0.37]	.096	**−0.24**	**[−0.43, −0.05]**	**.015**	**−0.37**	**[−0.56, −0.17]**	**<.001**
Purity	0.19	[−0.09, 0.47]	.183	−0.17	[−0.46, 0.12]	.249	**0.35**	**[0.07, 0.63]**	**.015**	**0.37**	**[0.09, 0.66]**	**.011**
Openness	**−0.25**	**[−0.43, −0.07]**	**.006**	0.04	[−0.15, 0.22]	.703	0.05	[−0.13, 0.23]	.575	0.10	[−0.08, 0.28]	.293
Emotional Stability	0.12	[−0.06, 0.30]	.186	−0.15	[−0.33, 0.03]	.105	**0.23**	**[0.05, 0.41]**	**.010**	0.14	[−0.04, 0.32]	.124
Extraversion	**0.22**	**[0.03, 0.41]**	**.023**	0.01	[−0.18, 0.21]	.881	**−0.58**	**[−0.77, −0.40]**	**<.001**	**−0.38**	**[−0.57, −0.19]**	**<.001**
Study 2												
Broad vs. specific	0.06	[−0.08, 0.20]	.420	**−0.17**	**[−0.32, −0.02]**	**.024**	**0.17**	**[0.02, 0.31]**	**.024**	−0.01	[−0.15, 0.14]	.899
Amplifying vs. curbing												
Amplifying	**−0.43**	**[−0.59, −0.26]**	**<.001**	0.14	[−0.03, 0.31]	.097	−0.08	[−0.25, 0.08]	.320	**0.34**	**[0.18, 0.51]**	**<.001**
Both	−0.13	[−0.30, 0.04]	.141	0.15	[−0.02, 0.33]	.089	0.04	[−0.14, 0.21]	.655	**0.22**	**[0.04, 0.39]**	**.016**
Trait content												
Agreeableness	−0.02	[−0.14, 0.11]	.800	0.02	[−0.11, 0.14]	.783	**−0.22**	**[−0.34, −0.09]**	**.001**	**0.16**	**[0.04, 0.29]**	**.011**
Conscientiousness	0.06	[−0.11, 0.23]	.470	−0.05	[−0.23, 0.12]	.552	0.00	[−0.17, 0.18]	.969	−0.03	[−0.20, 0.14]	.723
Honesty–Humility	0.01	[−0.19, 0.20]	.927	0.17	[−0.03, 0.37]	.101	−0.11	[−0.31, 0.09]	.270	−0.17	[−0.37, 0.03]	.092
Purity	**−0.30**	**[−0.60, 0.00]**	**.046**	0.09	[−0.21, 0.39]	.560	**0.49**	**[0.19, 0.79]**	**.002**	−0.10	[−0.40, 0.20]	.528
Openness	−0.01	[−0.28, 0.25]	.919	−0.08	[−0.35, 0.20]	.587	0.07	[−0.20, 0.34]	.633	0.15	[−0.12, 0.42]	.270
Emotional Stability	**0.35**	**[0.17, 0.52]**	**<.001**	**−0.20**	**[−0.38, −0.02]**	**.032**	0.10	[−0.08, 0.28]	.270	0.13	[−0.05, 0.31]	.159
Extraversion	−0.07	[−0.34, 0.20]	.612	0.09	[−0.19, 0.36]	.533	−0.20	[−0.47, 0.07]	.155	−0.18	[−0.45, 0.09]	.183

Note: Breadth (reference category: specific) and amplifying versus curbing (reference category: curbing) were dummy-coded; thus, effects are relative to the reference category. Trait domain variables (contrasting category: other) were effect-coded; thus, effects are relative to the unweighted grand mean across all trait domain categories. Dependent variables were standardized; categorical variables were not. Coefficients in **boldface** were detectable at $p < .05$.

TABLE 5 Multiple regression analysis predicting change motivation from all goal content and self-reported variables (Studies 1 and 2).

	Study 1			Study 2		
	β	95% CI	*p*	*β*	95% CI	*p*
Goal content variables						
Broad vs. specific	−0.03	[−0.19, 0.13]	.737	0.06	[−0.07, 0.19]	.375
Amplifying vs. curbing						
Amplifying	0.13	[−0.04, 0.31]	.141	−0.12	[−0.28, 0.03]	.106
Both	0.17	[−0.05, 0.38]	.131	0.03	[−0.13, 0.18]	.748
Trait content						
Agreeableness	**−0.25**	**[−0.37, −0.12]**	**<.001**	**−0.15**	**[−0.26, −0.04]**	**.009**
Conscientiousness	**−0.16**	**[−0.30, −0.01]**	**.035**	−0.08	[−0.23, 0.07]	.311
Honesty–Humility	0.13	[−0.05, 0.31]	.168	0.00	[−0.17, 0.18]	.976
Purity	0.24	[−0.02, 0.50]	.068	0.04	[−0.23, 0.31]	.782
Openness	0.12	[−0.05, 0.28]	.169	−0.06	[−0.30, 0.18]	.621
Emotional Stability	−0.02	[−0.18, 0.15]	.833	0.11	[−0.06, 0.27]	.197
Extraversion	−0.10	[−0.28, 0.08]	.267	0.08	[−0.16, 0.32]	.524
Goal perception variables						
Most motivated to change for: Self	**−0.13**	**[−0.26, −0.00]**	**.048**	−0.08	[−0.21, 0.05]	.240
Well-being consequences for self				**0.36**	**[0.30, 0.43]**	**<.001**
Positive consequences for self	**0.38**	**[0.31, 0.45]**	**<.001**			
Negative consequences for self	0.00	[−0.08, 0.07]	.951			
Well-being consequences for others				**0.17**	**[0.10, 0.23]**	**<.001**
Positive consequences for others	**0.14**	**[0.07, 0.20]**	**<.001**			
Negative consequences for others	0.01	[−0.05, 0.08]	.666			
Difficulty	0.01	[−0.06, 0.08]	.787	−0.01	[−0.07, 0.06]	.851
Controllability	**0.13**	**[0.07, 0.20]**	**<.001**	**0.13**	**[0.07, 0.19]**	**<.001**

Note: Breadth (reference category: specific) and amplifying versus curbing (reference category: curbing) were dummy-coded; thus, effects are relative to the reference category. Trait domain variables (contrasting category: other) were effect-coded; thus, effects are relative to the unweighted grand mean across all trait domain categories. Coefficients were standardized against both the dependent and predictor variables for continuous predictors, and only standardized against the dependent variable for categorical predictors. Coefficients in **boldface** were detectable at $p < .05$. In Study 1, well-being consequences were measured using separate items for positive and negative consequences; in Study 2, we used bipolar items to capture overall well-being consequences. See Table S2 for results of models with only coded variables or only self-reported variables.

have biased participants toward thinking about broad traits ("thing you want to change about *yourself*") and amplification-framed improvements ("in order to become *more* morally *good*"). The YourMorals sample may also be unique in at least two ways. First, these participants leaned substantially more liberal (50%) than conservative (16.32%), and supplemental analyses showed that, compared to conservatives, liberals were more interested in becoming more compassionate and less interested in becoming more pure (see Figure S1). Second, those who participate in moral psychology studies for fun (rather than for compensation) might care more about morality (or, at the very least, might be more interested in understanding and reflecting on their moral values) than the typical person. Thus, it seems important to know whether goals to become more open-minded would be similarly prevalent in a sample that is not characterized by an interest in moral psychology and moral self-exploration.

Two further potential sources of ambiguity are that we did not define the term "moral" to participants, and that the procedure assumed that participants wanted to make a moral improvement at all (when that may not have been the case; e.g., Schwitzgebel, 2019; Sun & Goodwin, 2020). Thus, it is possible that some of the potential improvements

that participants suggested they could make were neither "moral" nor "goals." Study 2 addresses each of these limitations and also assesses generalizability to self-codings (as opposed to expert codings).

3 | STUDY 2

3.1 | Method

3.1.1 | Participants

We used CloudResearch to recruit CloudResearch-approved Amazon Mechanical Turk participants who were based in the U.S. To avoid self-selection based on an interest in morality, the study was advertised as a "Personal change study" that would be about "the ways that people do or don't want to change." We preregistered a plan to stop data collection after recruiting 1200 participants who passed all three comprehension checks and submitted a completion code. The first two comprehension checks (which assessed participants' understanding of the task and our definition of morality) were designed to be very difficult to fail if participants were paying attention; we therefore automatically screened participants out of the survey if they failed these comprehension checks. The third comprehension check (described below) was more intensive; we therefore allowed as many attempts as were needed to pass. Due to a technical error in determining which participants contributed to the quota, we ended up with 1234 initial participants.

To ensure that the study only included participants who wanted to make a moral improvement, we asked if there was "anything that you WANT to change about yourself or your behavior in order to become more morally good or less morally bad?" As preregistered, 366 participants who said "No" completed a different version of the survey (see materials on the OSF for details) and were excluded from this study. Of the 868 participants who stated that they wanted to make a moral improvement, we excluded two participants who provided invalid responses (based on the same criteria as in Study 1) to the key open-ended question about their desired moral improvement (described below). After applying these preregistered exclusions, the final subset included 864 participants (442 women, 409 men, 3 preferred to self-describe, 9 not reported) who were between the ages of 18 and 84 (M = 40.74, SD = 12.61). Although this sample comprised a similar percentage of liberals (52.72%) as the YourMorals sample, it included a much larger share of conservatives (28.85%).

3.1.2 | Measures

Goal content variables

Participants who stated that they wanted to make a moral improvement were asked to "please describe ONE thing you want to change about yourself *or do differently* in order to become [more morally good *or less morally bad*; *less morally bad or* more morally good]" (differences from the Study 1 instructions are italicized here to facilitate comparison) using at least one complete sentence. We counterbalanced (between participants) whether "more morally good" or "less morally bad" appeared first. These modifications to the instructions were intended to counteract potential biases toward broad, amplification-focused improvements. After completing the goal perception variables (described in the next section), participants self-coded their own responses for the variables described below, in the following order.

Amplifying versus curbing. Participants selected whether the change involved stopping or decreasing something bad, starting or increasing something good, or both.

Trait content categories. Participants selected the trait category that best captured the change they just described. Specifically, we asked, "Which of the following traits best describes the change that you just described? By making this change, I would become..." We used the same coding scheme as in Study 1 (see Table 1), but only included the socially desirable poles (e.g., More Compassionate, Less Reactive) to simplify the response option set. Participants therefore self-coded their responses into one of 20 socially desirable changes or "Something else" (see Table 2 for example responses). Note that we intentionally asked participants to code the amplifying versus curbing variable (described above) before moving onto the trait content categories, to avoid the possibility that participants would be more likely to choose the amplifying option (i.e., "starting or increasing something good") because most of the trait content categories were framed in terms of obtaining "more" of a positive trait. For the regression analyses, we used the same broad trait groupings as in Study 1 (see Table 1).

To ensure that participants understood and applied our definitions of these trait categories (as much as was feasible within the context of a short survey, as opposed to years of training in personality research), before the task, participants read brief definitions of each trait. We asked participants to interpret these trait labels in line with our definitions (even if they had different ideas about what each of these terms mean). After reading these definitions, participants were asked to match up each trait with its definition as a final comprehension check. Because this task involved matching up 20 traits with 20 definitions (split up

into 3 blocks), and making fairly fine-grained distinctions between similar traits (e.g., responsibility vs. self-control), participants were allowed as many attempts as needed to pass this comprehension check. To remind participants of these definitions while they selected the trait category that best captured the change they just described, the response options included a brief definition of each trait.

Breadth of change goals. In Study 1, the expert codings of whether participants' responses reflected a goal to change a contextualized tendency or a broad trait were not particularly reliable ($\kappa = .53$). To improve on the measurement of this variable in Study 2, we asked participants three questions (instead of just one) about whether the change (1) involved one specific behavior or several different behaviors, (2) was primarily relevant to how they treat specific others or how they act in general, and (3) was primarily relevant to how they act in specific situations or how they act in general. If participants selected the first option for any of the three questions, we categorized their moral improvement goal as being contextualized. Otherwise, we categorized their moral improvement as being a broad trait change.

Goal perception variables

Motivation. Participants reported the extent to which they want to make this change (0 = *Not at all*, 10 = *A great deal*).

Beneficiaries. In response to the question, "For whose sake are you MOST motivated to make this change (i.e., whose interests do you most have at heart in wanting to make this change)?" participants selected one entity out of the following list: Myself, My family, My friends, My co-workers/company, My country/people in my country, All people, Non-human animals, The environment, Future generations, A specific group of people not mentioned here (please specify), Other (please specify).

Costs and benefits for the self and others. Participants selected the life/lives that would be most positively impacted if they made this change out of three mutually exclusive options: My own, Other person/people, Something else (e.g., animals).

Participants also reported their beliefs about the impact that making this change would have on their own well-being and, separately, the well-being of others (e.g., other people, animals; −4 = *Extremely negative impact*, 0 = *No impact*, 4 = *Extremely positive impact*).

Controllability. Participants reported the extent to which they think making this change is within their control (0 = *Not at all within my control*, 10 = *Extremely within my control*).

Difficulty. Participants reported how difficult it would be for them to make the change (0 = *Not difficult at all*, 10 = *Extremely difficult*).

3.1.3 | Data analyses

As in Study 1, analyses were conducted in R (R Core Team, 2022), and we used the same dummy- and effect-coding strategies.

3.2 | Results

3.2.1 | How do people conceptualize moral improvement?

In contrast to Study 1, participants' self-codings showed that they were overwhelmingly interested in improving in specific, contextualized ways (68.4%) rather than in improving broad character traits (31.6%). As in Study 1, people more often reported desires to improve or amplify positive tendencies (40.97%), rather than to curb negative ones (28.47%; e.g., "be less judgemental") or to both curb negative qualities and amplify positive ones (30.56%; e.g., "Increase patience and decrease reactivity"). However, this difference was relatively less pronounced than in Study 1.

As in Study 1, participants reported a diverse range of moral improvement goals that spanned all 20 trait categories, as well as traits that were not included in the coding scheme (see Figure 1). Consistent with the results of the expert codings in Study 1, participants' self-codings in Study 2 showed that goals to become more compassionate were the most frequently desired moral improvements (27.66%). Similarly, goals to become less reactive (9.95%) and more honest (8.33%) were once again among the top five moral improvement goals, whereas goals to become more fair (2.55%), loyal (2.66%), humble (2.08%), and grateful (1.74%) were once again surprisingly rare (considering that these are often considered to be moral virtues). Compared to Study 1, goals to become more open-minded (difference = −8.65%), productive (difference = −7.11%), and brave (difference = −4.40%) were relatively less prevalent, whereas goals to become more respectful (difference = 5.92%) and more self-controlled (difference = 3.62%) were relatively more prevalent.

3.2.2 | What do people believe about the process and consequences of moral improvement?

Next, we explored people's beliefs about the process and consequences of moral improvement. As shown in Table 6,

TABLE 6 Descriptive statistics and correlations among continuous variables (Study 2).

Variable	*M*	*SD*	1.	2.	3.	4.
1. Change motivation	8.17	1.66				
2. Consequences for own well-being	2.51	1.27	.45			
3. Consequences for others' well-being	2.47	1.31	.31	.36		
4. Difficulty	6.75	2.44	−.05	−.04	−.08	
5. Controllability	8.21	1.98	.21	.15	.16	−.29

Note: $rs \geq |.08|$ were detectable at $p < .05$. The change motivation, difficulty, and controllability variables were measured on 0–10 scales, and the well-being consequences variables were measured on −4 (very negative impact) to 4 (very positive impact) scales.

although people thought that making their desired moral change would be quite difficult, they also thought that it was largely within their control. Replicating the results of Study 1, when asked whose life/lives would be most positively impacted, almost half of participants selected "Me" (49.42%), followed by "Other person/people" (46.88%), and "Something else" (3.70%). Once again, compassion was an exception to this tendency (see Figure S2). People also believed that moral improvements would have moderately to very positive consequences for both their own well-being and the well-being of others (see Table 6). These findings once again suggest that people overwhelmingly forecast substantial benefits of becoming more moral, for both themselves and for others.

Which changes did people think would be more difficult, more controllable, and result in more favorable consequences for themselves and for others? The following findings from Study 1 replicated in Study 2. First, people believed that it is less difficult to amplify positive qualities than to curb negative qualities. Second, people believed that improvements to broad traits (compared with specific or contextualized improvements) and purity-related improvements would result in more favorable well-being consequences for themselves (see Table 4), whereas improving their agreeableness would result in worse well-being consequences for themselves (each compared with the unweighted grand mean). Third, people also believed that agreeableness-related improvements would result in better well-being consequences for others (compared with the unweighted grand mean).

However, we found a different pattern of results for the trait content of the improvements that people believed would be more or less difficult to make (see Table 4). Unlike Study 1, people believed that purity-related changes would be less difficult to make and that emotional stability-related changes would be more difficult to make (compared with the unweighted grand mean). In addition, whereas there were no statistically detectable differences in the perceived controllability of various kinds of moral improvement (e.g., broad vs. specific) in Study 1, in Study 2, people reported that broad trait changes (vs. contextualized changes) and emotional stability changes (vs. the unweighted grand mean) would be less within their control. In Study 2, people also believed that amplifying positive tendencies or both amplifying positive tendencies and curbing negative tendencies would have better well-being consequences for others (compared with only curbing negative tendencies).

3.2.3 | What motivates people to be more moral?

Finally, we turned to the goal of understanding what motivates people to be more moral. As in Study 1, when asked to indicate for whose sake they were most motivated to make the moral change they mentioned, almost half of the participants chose themselves (40.05%). This was followed by all people (20.02%), family (19.56%), friends (4.05%), their country/people in their country (3.82%), a specific group of people not mentioned here (2.89%), the environment (2.66%), other (2.43%), future generations (1.97%), non-human animals (1.39%), and their co-workers/company (1.16%).

Next, we predicted motivation to enact the desired change from the self-coded content of the goals, as well as self-reported beliefs about the consequences, difficulty, and controllability of the change. We did so separately for the set of self-coded variables ($R^2 = .03$), separately for the set of self-reported variables ($R^2 = .25$), and finally, with all variables included in the model ($R^2 = .26$). As in Study 1, the final model showed that people who reported agreeableness-related improvements and who believed that the change was less within their control were less motivated to make it (see Table 5). In addition, people who believed that a change would result in better well-being consequences for both the self and for others were more motivated to make it. As in Study 1, a nonpreregistered analysis showed that perceived well-being consequences for the self were a stronger predictor than perceived

well-being consequences for others, $\Delta\chi^2$ ($df = 1$) = 8.94, $p < .001$.

Unlike Study 1, there were no detectable differences in motivation to undertake the change for those who were most motivated to make the change for their own (vs. others') sake. Nor did we replicate the finding that goals to become more conscientious were less motivating.

3.3 | Discussion

In sum, Study 2 replicated most of the key findings from Study 1 while addressing concerns about potential wording biases, methodological ambiguities, and sampling biases. Specifically, we found that people more frequently self-categorize their moral improvement goals in terms of amplifying positive qualities than curbing negative qualities. However, this difference was not as dramatic as in Study 1, perhaps due to the adjustments to the wording of the instructions. In addition, as in Study 1, the goal to become more compassionate was the most prevalent. We also replicated the findings that people believe that they would be the primary beneficiaries of their moral improvement, and that they were more strongly motivated by the perceived well-being consequences for themselves than the perceived well-being consequences for others.

One major difference compared with Study 1 is that participants' self-codings in Study 2 suggested that their moral improvements tended to be about a specific behavior or were contextualized within a specific context or relationship (rather than reflecting a broad trait change, as Study 1 would suggest). This discrepancy could be due to two methodological improvements: the adjustment to the instruction wordings (which, in Study 1, may have been biased towards broad trait changes) and the use of three questions (about specificity in relation to behaviors, relationships, and situations) instead of one. Alternatively, it could be the case that people tended to imprecisely express their desired moral improvements in general terms, even when they actually had a specific behavior, relationship, or situation in mind. This illustrates the value of using self-codings (which are based on additional contextual information) to complement researcher codings (which are based more literally on what participants express). All things considered, we believe that the conclusion from Study 2—that moral improvement goals more often tend to be contextualized, rather than broad—is more likely to be true.

On the other hand, very few of the results that used the trait domains as predictors were consistent between the two studies. We believe that this is because participants were unable to reliably self-categorize their goals based on the admittedly complex trait taxonomy we provided them with in Study 2, despite our efforts to align participants' definitions with ours. Indeed, for a random sample of 200 responses, the agreement between self-codings and expert codings (provided by the first and second authors, with the third author resolving disagreements) was $\kappa = .42$ (50.5% agreement) for the 21-trait category taxonomy and $\kappa = .44$ (58.5% agreement) for the broad trait domain. Nevertheless, several of the same trait categories that were among the most frequently mentioned in Study 1—goals to be more compassionate, less reactive, and more honest—also emerged in Study 2. This suggests that the self-codings were at least reliable enough for the purposes of estimating the relative representation of various traits that are reflected by people's moral improvement goals. However, the low reliability of these self-codings—as well as differences in populations and the possibility of false positives—might partially explain why some of the trait predictor findings that emerged in Study 1 (e.g., that people were relatively less motivated to become more conscientious) did not replicate in Study 2.

4 | GENERAL DISCUSSION

How do people conceptualize moral improvement, and when do they feel more motivated to undertake such improvements? Table 7 shows a summary of the key findings. Across two large studies that featured two very different samples (YourMorals.org vs. CloudResearch) and methods (expert codings vs. self-codings), we find that people conceptualize moral improvement in diverse ways. A wide range of traits was represented, but the most common goal was to become more compassionate, followed by goals to become less reactive and more honest. People were somewhat more inclined to conceptualize moral improvement in terms of starting or increasing positive tendencies (rather than stopping or decreasing negative ones). People were less motivated to make agreeableness-related improvements. They were more motivated to make improvements that they perceived as being within their control. They were also more motivated to make improvements that they believed would benefit themselves, and to a less pronounced extent, others.

4.1 | Diverse concepts of moral improvement

Moral relevance norms from past studies show that compassion, respectfulness, honesty, fairness, bravery, and loyalty are among the most prototypical moral traits (Goodwin et al., 2014; Strohminger & Nichols, 2015; Sun & Goodwin, 2020). Do people think of moral

TABLE 7 Summary of key findings.

Research question	Study 1 (expert codings)	Study 2 (self-codings)
How do people typically conceptualize moral improvement?		
Breadth	Broad trait changes (81.24%) Specific/contextualized changes (18.76%)	Specific/contextualized changes (68.4%) Broad trait changes (31.6%)
Amplifying vs. curbing	**Amplifying (66.35%)** Curbing (16.88%) Both (16.56%)	**Amplifying (40.97%)** Both (30.56%) Curbing (28.47%)
Most frequently mentioned five trait changes	↑ **Compassion**, ↑ Open-mindedness, ↑ Productiveness, ↓ **Reactivity**, ↑ **Honesty**	↑ **Compassion**, ↓ **Reactivity**, ↑ **Honesty**, ↑ Respectfulness, ↑ Self-control
What do people believe about the process and consequences of moral improvement?		
Difficulty	**Amplifying (−)**, Agreeableness (−), Openness (−), Extraversion (+)	**Amplifying (−)**, Purity (−), Emotional Stability (+)
Controllability	None	Broad (−), Emotional Stability (−)
Positive consequences for self	**Broad (+)**, **Agreeableness (−)**, Conscientiousness (+), Honesty–Humility (−), **Purity (+)**, Emotional Stability (+), Extraversion (−)	**Broad (+)**, **Agreeableness (−)**, **Purity (+)**
Positive consequences for others	**Agreeableness (+)**, Honesty–Humility (−), Purity (+), Extraversion (−)	Amplifying (+), Both Amplifying and Curbing (+), **Agreeableness (+)**
What motivates people to be more moral?		
Most motivated to change for...	**Myself (48.30%)**, immediate family members (16.02%), people on all continents (7.49%); all other categories <7% each	**Myself (40.05%)**, all people (20.02%), family (19.56%); all other categories <5% each
Predictors of change motivation	**Agreeableness (−)**, Conscientiousness (−), Most motivated to change for: Self (−), **Positive consequences for self (+)**, **Positive consequences for others (+)**, **Controllability (+)**	**Agreeableness (−)**, **Well-being consequences for self (+)**, **Well-being consequences for others (+)**, **Controllability (+)**

Note: Results that are consistent across both studies—and therefore, more robust to differences in populations, instructions, and measurement approaches—are highlighted in **bold**. (+) and (−) denote positive and negative regression coefficients, respectively.

improvements in terms of these prototypical moral virtues? Demonstrating the diversity of people's concepts of moral improvement, moral improvement goals spanned more than 20 trait categories in both studies. Goals to become more compassionate, more honest, and more respectful, however, were among the five most frequently mentioned moral improvements in at least one of the two samples. In contrast, goals to become more loyal and more fair were mentioned less than 5% of the time in both samples.

It is particularly striking that goals to become more compassionate were, by far, the most frequently mentioned moral improvements. Indeed, in Study 2, compassion was mentioned almost three times more frequently than the second most common trait category (goals to become less reactive). Why might this be the case? There are at least two nonmutually exclusive interpretations for why some trait categories (e.g., becoming more compassionate) were more frequently mentioned than others (e.g., becoming more honest, fair, or loyal). First, perhaps these are the traits that are most relevant to the moral situations people encounter in everyday life. Providing some support for the relevance hypothesis, an experience sampling study of moral events in daily life suggests that Care/Harm (which corresponds most closely with the trait of compassion) was the most frequently mentioned moral domain

(Hofmann et al., 2014). Thus, perhaps people most frequently mentioned goals to become more compassionate in part because this is the most relevant moral domain in everyday life.

A second possibility is that people tend to perceive more opportunities for improvement on the traits that they more frequently mentioned. Compassion may be distinct from traits such as honesty, fairness, and loyalty, in at least two ways. First, judgments of honesty, fairness, and loyalty may be based more on behavior than are judgments of compassion. That is, whereas honest behavior (vs. "honest thoughts" or "honest feelings") may be the primary basis for judging oneself as being an honest person, people might consider themselves to be truly compassionate only if they think, feel, and act compassionately. Second, honesty, fairness, and loyalty may have consequences for a smaller set of potential targets compared with compassion. For example, one could be loyal to a romantic partner, friends, organization, or nation, but not to everyone in the world. In contrast, the scope of one's compassion could be extended beyond close others to not only strangers in one's community (e.g., homeless people), but also those in foreign countries (e.g., people who are dying from preventable diseases), nonhuman animals, and future generations (Singer, 1981). For these reasons, it may be easier to perfectly fulfill one's duties to be honest (by being completely truthful), fair (by being entirely unbiased), and loyal (by never betraying others), whereas the opportunities for becoming more compassionate seem almost endless by comparison (Trammell, 1975; see also Trafimow & Trafimow, 1999).

Interestingly, improvements to several traits that are not inherently moral traits—decreased reactivity and increased open-mindedness, productiveness, and self-control—were also among the most frequently mentioned moral improvements. These results suggest that even if certain traits are not inherently or specifically moral (e.g., self-control can be deployed toward moral, immoral, or morally neutral ends; Hofmann et al., 2018; Landy et al., 2016), people may still see such traits as being additional means to achieving moral ends.

4.2 | Self-interest is a key motivation for moral improvement

What motivates people to be more moral? From the perspective that the function of morality is to suppress selfishness for the benefit of others (Haidt & Kesebir, 2010; Wolf, 1982), we might expect people to believe that moral improvements would primarily benefit others (rather than themselves). By a similar logic, people should also primarily want to be more moral for the sake of others (rather than for their own sake).

Surprisingly, however, this was not overwhelmingly the case. Instead, across both studies, participants were approximately equally split between those who believed that others would benefit the most and those who believed that they themselves would benefit the most (with the exception of compassion; see Figure S2). The finding that people perceive some personal benefits to becoming more moral has been demonstrated in recent research (Sun & Berman, in prep). In light of evidence that moral people tend to be happier (Sun et al., in prep) and that the presence of moral struggles predicts symptoms of depression and anxiety (Exline et al., 2014), such beliefs might also be somewhat accurate. However, it is unclear why people believe that becoming more moral would benefit themselves more than it would others. Speculatively, one possibility is that people can more vividly imagine the impacts of their own actions on their own well-being, whereas they are much more uncertain about how their actions would affect others—especially when the impacts might be spread across many beneficiaries.

However, it is also possible that this finding only applies to self-selected moral improvements, rather than the universe of all possible moral improvements. That is, when asked what they could do to become more moral, people might more readily think of improvements that would improve their own well-being to a greater extent than the well-being of others. But, if we were to ask people to predict who would benefit the most from various moral improvements that were selected by researchers, people may be less likely to believe that it would be themselves. Future research should systematically study people's evaluations of how various moral improvements would impact their own and others' well-being.

Similarly, when explicitly asked for whose sake they were most motivated to make their moral improvement, almost half of the participants admitted that they were most motivated to change for their *own* sake (rather than for the sake of others). However, when predicting motivation from both the expected well-being consequences for the self and the well-being consequences for others, we found that people's perceptions of personal well-being consequences were a significantly stronger predictor in both studies. In other words, if anything, people are relatively more motivated to make moral improvements for their own sake than for the sake of others. This is consistent with the findings of another study that examined people's interest in changing a variety of moral and nonmoral traits, and showed that people are particularly interested in improving the traits that they believed would make them relatively happier (Sun & Berman, in prep). It was striking that in the present study, personal fulfillment remained the most important motivator of personal improvement even exclusively in the moral domain.

4.3 | Limitations and future directions

One limitation is that we only asked people about a moral improvement that they could make, rather than one that they were actively trying to make. Thus, it may be more accurate to describe these as "ideas" for moral improvement, rather than as active moral goals. In addition, hypothetical motivation may not translate into actual behavior. To address these limitations, future research could include only people who state that they are currently trying to improve in a moral way, and track their efforts and success at moral improvement over time. In the meanwhile, we believe that this study nevertheless provides valuable insight into how ordinary people envision moral improvement.

A further open question is whether and how moral improvements tend to be contextualized. Studies 1 and 2 provided conflicting answers to this question. Researcher codings of responses in Study 1 suggested that people describe their desired moral improvements in general terms, but participants' self-categorizations in Study 2 suggested that they generally had specific behaviors, relationships, or situations in mind. Future research should aim to understand what these specific behaviors, relationships, or situations are, and their implications for moral motivation.

Finally, there are two important constraints on generality. First, all goals were self-selected. This was an intentional design decision, as an idiographic approach was necessary to understand how people spontaneously envision moral improvement for themselves. However, it is an open question whether people would have similar beliefs about the controllability, difficulty, and well-being consequences of moral improvements that are imposed or encouraged by others, and whether the predictors of moral motivation that we found in this study would generalize to non-self-selected moral improvements as well (e.g., in the context of character education).

A second constraint on generality is that the vast majority of our participants resided in WEIRD countries (Henrich et al., 2010). Thus, our findings primarily provide insight into WEIRD concepts of moral improvement. Moreover, even within WEIRD contexts, culture matters. For example, we found that YourMorals participants (Study 1) more frequently emphasized goals to be more open-minded and productive, whereas CloudResearch participants (Study 2) more frequently emphasized goals to be more respectful. In addition, supplemental analyses showed that in both samples, political conservatives placed more emphasis on goals to be more pure or spiritual compared to liberals, whereas liberals tended to place more emphasis on goals to be more compassionate than did conservatives (although this was only marginally significant in Study 2).

5 | CONCLUSIONS

Moral improvement has long been the prerogative of religion and philosophy, but we propose that there is also value in better understanding people's personal moral strivings. Here, we show that people consider a diverse range of trait domains to be relevant to moral improvement, but place particular emphasis on becoming more compassionate, more honest, and less reactive. Surprisingly, people believe that their moral improvements would benefit themselves more than others. Finally, the perceived benefits to others, and—to an even greater extent—to the self, both motivate moral improvement. These findings provide rich descriptive insights into how ordinary people want to be more moral, and the role of well-being considerations in motivating moral improvement.

AUTHOR CONTRIBUTIONS

All authors contributed to conceptualization, methodology, and reviewing and editing the manuscript. J.S. and J.W. developed the coding scheme, and J.S., J.W., H.M.W., and P.M. contributed to the expert codings. P.M. collected the data for Study 1, and J.S. collected the data for Study 2. J.S. curated, analyzed, and visualized the data, supervised research assistants, and wrote the original draft.

ACKNOWLEDGMENTS

We thank Sophie Faircloth, Sarah Swartz, Laura Blanco, Joshua Honig-Frand, and Stephanie Hanchak for their research assistance with coding responses in Study 1 and conducting data quality checks for Study 2.

ORCID

Jessie Sun https://orcid.org/0000-0001-6764-0721
Joshua Wilt https://orcid.org/0000-0001-7059-6389
Peter Meindl https://orcid.org/0000-0002-2686-3723
Hanne M. Watkins https://orcid.org/0000-0002-8429-4661
Geoffrey P. Goodwin https://orcid.org/0000-0002-3682-0444

REFERENCES

Althof, W., & Berkowitz, M. W. (2006). Moral education and character education: Their relationship and roles in citizenship education. *Journal of Moral Education*, *35*(4), 495–518. https://doi.org/10.1080/03057240601012204

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

Crimston, C. R., Bain, P. G., Hornsey, M. J., & Bastian, B. (2016). Moral expansiveness: Examining variability in the extension of

the moral world. *Journal of Personality and Social Psychology*, *111*, 636–653. https://doi.org/10.1037/pspp0000086

Elliot, A. J., & Thrash, T. M. (2002). Approach-avoidance motivation in personality: Approach and avoidance temperaments and goals. *Journal of Personality and Social Psychology*, *82*(5), 804–818. https://doi.org/10.1037/0022-3514.82.5.804

Exline, J. J., Pargament, K. I., Grubbs, J. B., & Yali, A. M. (2014). The religious and spiritual struggles scale: Development and initial validation. *Psychology of Religion and Spirituality*, *6*, 208–222. https://doi.org/10.1037/a0036465

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

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. In *Advances in experimental social psychology* (Vol. *47*, pp. 55–130). Academic Press. https://doi.org/10.1016/B978-0-12-407236-7.00002-4

Grudem, W. (2018). *Christian ethics: An introduction to biblical moral reasoning*. Crossway.

Haidt, J., & Kesebir, S. (2010). Morality. In *Handbook of social psychology* (Vol. *2*, 5th ed., pp. 797–832). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470561119.socpsy002022

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

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*, 1340–1343. https://doi.org/10.1126/science.1251560

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

Janoff-Bulman, R., Sheikh, S., & Hepp, S. (2009). Proscriptive versus prescriptive morality: Two faces of moral regulation. *Journal of Personality and Social Psychology*, *96*, 521–537. https://doi.org/10.1037/a0013779

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

Lavy, S. (2020). A review of character strengths interventions in twenty-first-century schools: Their importance and how they can be fostered. *Applied Research in Quality of Life*, *15*(2), 573–596. https://doi.org/10.1007/s11482-018-9700-6

Lee, K., & Ashton, M. C. (2018). Psychometric properties of the HEXACO-100. *Assessment*, *25*(5), 543–556. https://doi.org/10.1177/1073191116659134

McAdams, D. P. (1995). What do we know when we know a person? *Journal of Personality*, *63*(3), 365–396. https://doi.org/10.1111/j.1467-6494.1995.tb00500.x

Miller, T. J. (2022). Assessing the desire to change personality across methods. *Journal of Personality Assessment*, *104*(4), 447–457. https://doi.org/10.1080/00223891.2021.1955695

Persson, I., & Savulescu, J. (2012). *Unfit for the future: The need for moral enhancement*. Oxford University Press.

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

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.Rproject.org/

Schwitzgebel, E. (2019). Aiming for moral mediocrity. *Res Philosophica*, *96*(3), 347–368. https://doi.org/10.11612/resphil.1806

Singer, P. (1981). *The expanding circle: Ethics and sociobiology*. Farrar, Straus & Giroux.

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*(1), 117–143. https://doi.org/10.1037/pspp0000096

Specker, J., Focquaert, F., Raus, K., Sterckx, S., & Schermer, M. (2014). The ethical desirability of moral bioenhancement: A review of reasons. *BMC Medical Ethics*, *15*(1), 67. https://doi.org/10.1186/1472-6939-15-67

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

Sun, J., & Berman, J. Z. (in prep). *People prioritize personal fulfilment over morality in the context of self-improvement*. Manuscript in preparation. https://psyarxiv.com/gd5pw

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., Wu, W., & Goodwin, G. P. (in prep). *Moral people tend to be happier*. Manuscript in preparation. https://psyarxiv.com/sd8t4/

Telushkin, J. (2011). *The book of Jewish values: A day-by-day guide to ethical living*. Bell Tower.

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

Trammell, R. L. (1975). Saving life and taking life. *The Journal of Philosophy*, *72*(5), 131–137. https://doi.org/10.2307/2024750

Wilburn, B. (2010). *Moral cultivation: Essays on the development of character*. Lexington Books.

Wolf, S. (1982). Moral saints. *The Journal of Philosophy*, *79*(8), 419–439. https://doi.org/10.2307/2026228

SUPPORTING INFORMATION

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

How to cite this article: Sun, J., Wilt, J., Meindl, P., Watkins, H. M., & Goodwin, G. P. (2023). How and why people want to be more moral. *Journal of Personality*, *00*, 1–19. https://doi.org/10.1111/jopy.12812

1. Developing and Refining the Trait Content Coding Scheme

Given the exploratory goals of the project, we split the Study 1 dataset into exploratory (n = 487) and confirmatory subsets (n = 1,070). We used the exploratory subset to develop the preregistered analysis plan and refine the coding scheme that we used for the confirmatory subset (reported in the main text). After excluding invalid responses (based on the same criteria as in the main text) and the first 50 responses (which we used to calibrate the expert coders' interpretations of the coding scheme), the exploratory subset comprised 332 valid responses. Below, we report the key findings that informed our final coding scheme.

As a starting point, we included the 15 facets of the Big Five Inventory–2 (BFI-2; Soto & John, 2017). We then added patience and 11 traits from three additional taxonomies—the 24 VIA Character Strengths (Peterson & Seligman, 2004), Moral Foundations Theory (Graham et al., 2013), and facets of the HEXACO Honesty-Humility domain (Lee & Ashton, 2018)—to include traits that were not represented by the BFI-2 facets. Therefore, the coding scheme initially included 27 trait categories (see Table S1), plus an "Other" category (for any traits that could not be captured by this coding scheme). Inter-rater agreement was $\kappa = .69$, and a third expert resolved disagreements (27.58%). However, as shown in Table S1, none of the participants' moral goals reflected the categories of authority, humor, and creativity, and only one moral goal reflected the aesthetic sensitivity and emotional volatility categories, respectively. In light of these low base rates, we chose to drop the authority, humor, and aesthetic sensitivity categories from the final coding scheme (see Table 1). We also noticed conceptual or practical similarities (i.e., similar goals) between some of the categories (see Table S2). We therefore decided to combine perspective, intellectual curiosity, and creativity into a more inclusive open-mindedness category, to combine emotional volatility and low patience into a

reactivity category, and to expand the definition of bravery to include assertiveness. For completeness, definitions of the trait categories that were dropped or combined, along with examples (where available), are shown in Table S2.

Table S1

Frequencies of Goals Within Each Trait Content Category in the Exploratory Coding Scheme

Trait	% Goals to Increase	% Goals to Decrease	Included in Final Coding Scheme?
Compassion	23.19	0	Yes
Perspective	8.13	0	Yes, within Open-mindedness
Productiveness	7.23	0	Yes
Patience	6.33	0	Yes, within Reactivity
Trust	6.02	0.6	Yes
Bravery	5.42	0	Yes, but with the addition of Assertiveness
Purity/Spirituality	5.42	0	Yes
Honesty	5.12	0	Yes
Responsibility	4.52	0	Yes
Respectfulness	3.92	0	Yes
Self-control	3.61	0	Yes
Anxiety	0	2.71	Yes
Sociability	2.71	0	Yes
Depression	0	1.81	Yes
Intellectual curiosity	1.51	0	Yes, within Open-mindedness
Assertiveness	1.2	0	Yes, within Bravery
Humility	1.2	0	Yes
Loyalty	1.2	0	Yes
Energy level	0.9	0	Yes
Gratitude	0.9	0	Yes
Organization	0.9	0	Yes
Fairness	0.6	0	Yes
Emotional volatility	0	0.3	Yes, within Reactivity
Aesthetic sensitivity	0.3	0	No
Authority	0	0	No
Creativity	0	0	Yes, within Open-mindedness
Humor	0	0	No

Note. Trait categories are ordered from most frequently mentioned to least frequently mentioned. 4.44% of the responses were coded as "Other."

Table S2

Definitions and Examples of Goals Within Trait Categories that were Dropped or Combined in the Final Coding Scheme

Trait	Definition	Example
Perspective	Analytically evaluating ideas; seeing the bigger picture and multiple sides of an issue; intellectual humility.	I would like to try to understand and accept other people's political viewpoints.
Intellectual curiosity	Intellectual interests and enjoyment of thinking.	To perpetually become more understanding of myself and other people through continuous study including the works of great philosophers, theologians, and contemporary researchers.
Creative imagination	Creativity and originality.	N/A
Assertiveness	Willingness to express personal opinions and goals in social situations.	...more assertive with my opinions in order to stay true to my moral values.
Bravery	Acting on convictions and facing fears; includes physical, psychological, and moral bravery.	To more frequently speak up in real time when I see bad things happening in the world around me.
Emotional volatility	The tendency to experience volatile mood swings.	Less prone to overreaction - being more centered and thoughtful in my conduct.
Patience	The propensity to wait calmly in the face of frustration or adversity; not being easily angered or irritated.	I want to be slower to react with anger when someone expresses an opinion I disagree with or dislike.
Authority	Obedience and respect for authority and tradition.	N/A
Aesthetic sensitivity	Being interested in and valuing art, music, literature, and beauty.	I would like to change professions into a career where I can enact positive change in the world. I'm in the military right now, and no longer believe in the mission…I want to build beautiful furniture out of wood for a living and make sure that people can have beautiful, permanent things in their homes.
Humor	The ability to make other people smile or laugh, offer the lighter side to others, and have a composed and cheerful view on adversity.	N/A

Note. N/A = no example available. Ellipses indicate that the examples have been truncated. Examples have been corrected for spelling and grammar.

2. Predicting Change Motivation Separately from Goal Content or Goal Perception Variables

Table S3

Multiple Regression Analyses Predicting Change Motivation Separately from Goal Content or Goal Perception Variables

	Study 1			Study 2		
	β	95% CI	*p*	*β*	95% CI	*p*
Model with only goal content variables						
Broad vs. specific	0.09	[–0.08, 0.26]	.301	0.10	[–0.05, 0.24]	.201
Amplifying vs. curbing						
Amplifying	0.18	[–0.02, 0.37]	.071	–0.07	[–0.24, 0.10]	.419
Both	0.18	[–0.05, 0.42]	.126	0.10	[–0.07, 0.28]	.252
Trait content						
Agreeableness	**–0.36**	**[–0.49, –0.22]**	**< .001**	**–0.19**	**[–0.31, –0.06]**	**.004**
Conscientiousness	–0.06	[–0.22, 0.10]	.455	–0.09	[–0.27, 0.08]	.297
Honesty-Humility	0.11	[–0.08, 0.31]	.265	–0.05	[–0.25, 0.15]	.634
Purity	**0.38**	**[0.10, 0.67]**	**.009**	0.21	[–0.10, 0.51]	.179
Openness	0.06	[–0.12, 0.24]	.514	–0.02	[–0.29, 0.25]	.885
Emotional Stability	0.03	[–0.15, 0.21]	.760	0.13	[–0.05, 0.31]	.154
Extraversion	**–0.22**	**[–0.41, –0.03]**	**.026**	–0.01	[–0.28, 0.27]	.956
Model with only goal perception variables						
Most motivated to change for: Self	–0.13	[–0.26, +0.00]	.055	–0.05	[–0.18, 0.08]	.445
Well-being consequences for self				**0.39**	**[0.32, 0.45]**	**< .001**
Positive consequences for self	**0.39**	**[0.32, 0.46]**	**< .001**			
Negative consequences for self	–0.01	[–0.08, 0.06]	.785			
Well-being consequences for others				**0.15**	**[0.08, 0.21]**	**< .001**
Positive consequences for others	**0.14**	**[0.07, 0.20]**	**< .001**			
Negative consequences for others	0.03	[–0.04, 0.09]	.433			
Difficulty	0.02	[–0.05, 0.08]	.646	0.02	[–0.05, 0.08]	.628
Controllability	**0.14**	**[0.08, 0.20]**	**< .001**	**0.13**	**[0.07, 0.19]**	**< .001**

Note. Breadth (reference category: specific) and amplifying vs. curbing (reference category: curbing) were dummy-coded; thus, effects are relative to the reference category. Trait domain variables (contrasting category: other) were effect-coded; thus, effects are relative to the unweighted grand mean across all trait domain categories. Coefficients were standardized against both the dependent and predictor variables for continuous predictors, and only standardized against the dependent variable for categorical predictors. Coefficients in **boldface** were detectable at $p < .05$. In Study 1, well-being consequences were measured using separate items for positive and negative consequences; in Study 2, we used bipolar items to capture overall well-being consequences. See Table 5 for results of the final model in which all coded variables and self-reported variables were simultaneously entered as predictors.

3. Political Orientation Analyses

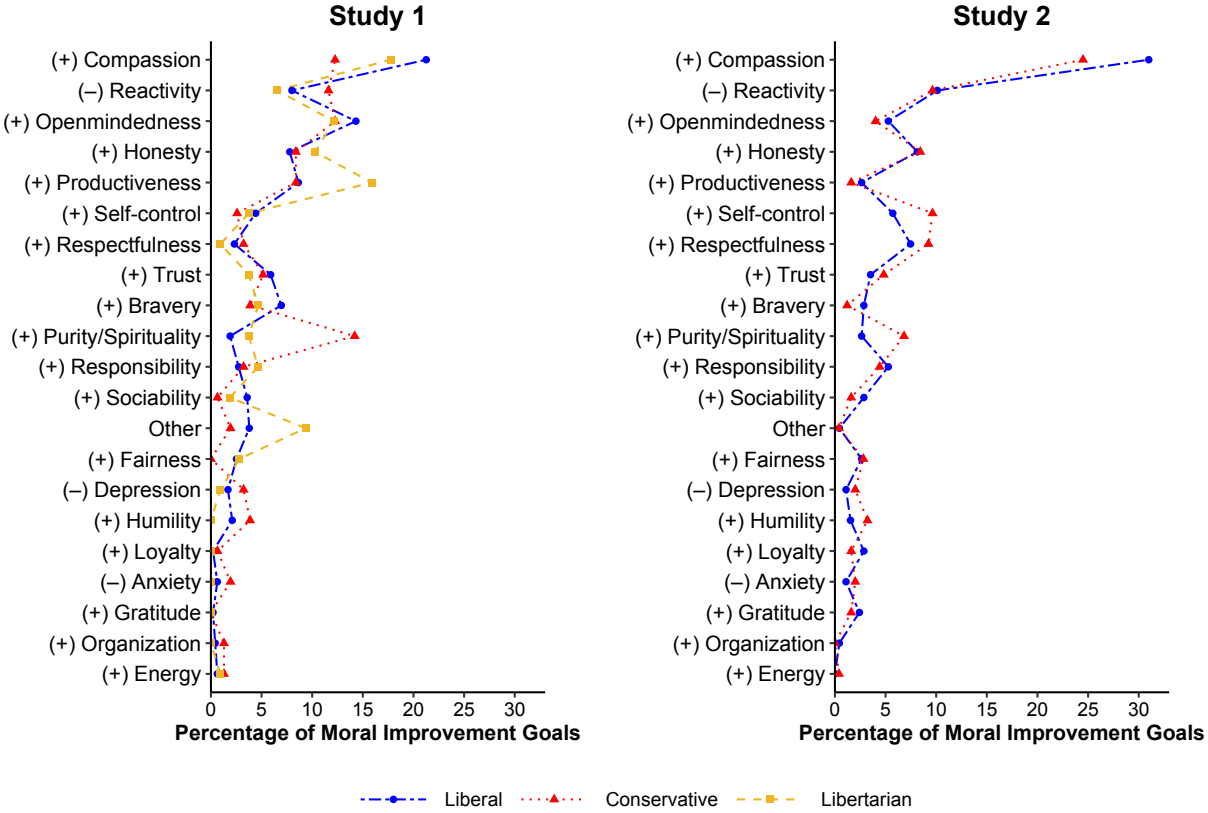
To examine the generalizability of the trait content of the moral improvement goals across political orientations, we compared the breakdowns for liberals, conservatives, and libertarians. Participants indicated their political orientation using 10 options. For these analyses, those who indicated being Slightly Liberal, Liberal, or Very Liberal were categorized as being Liberal (n_{Study1} = 475 [50%]; n_{Study2} = 455 [52.72%]), those who indicated being Slightly Conservative, Conservative, or Very Conservative were categorized as being Conservative (n_{Study1} = 155 [16.32%]; n_{Study2} = 249 [28.85%]), those who were Libertarian were categorized as being Libertarian (n_{Study1} = 107 [11.26%]; n_{Study2} = 17 [1.97%]), and those who selected the Moderate, Don't know/not political, or Other options (n_{Study1} = 213 [22.42%]; n_{Study2} = 142 [16.45%]) were excluded from this analysis.

Figure S1 shows the percentage of moral improvement goals separately for the three groups in Study 1, and for liberals vs. conservatives in Study 2 (because there were only 17 libertarians in Study 2, we deviated from our preregistered plan to include libertarians in this analysis). Overall, preregistered comparisons revealed very few differences between liberals and conservatives. Tests of the equality of proportions showed that liberals more frequently mentioned goals to become more compassionate than conservatives in Study 1 (χ^2 = 5.58, p = .018), but this difference was only marginally significant in Study 2 (χ^2 = 3.00, p = .083). Across both studies, however, conservatives more frequently mentioned goals to become more pure/spiritual, compared to liberals (Study 1: χ^2 = 35.20, p < .001; Study 2: χ^2 = 6.13, p = .013). Although past work suggests that liberals tend to be more open-minded than conservatives (Duckitt & Sibley, 2010), we found no evidence in the extent to which liberals and conservatives

reported goals to become more open-minded (Study 1: $\chi^2 = 0.26$, $p = .61$; Study 2: $\chi^2 = 0.31$, $p = .575$).

Figure S1

Trait Categorizations of the Most Frequently Mentioned Moral Improvement Goals, Separately By Political Orientation



Note. Goals to change in a socially undesirable direction were recoded as "Other" in Study 1. Goals are ordered from most to least frequent across the entire sample (including those who were excluded from the political orientation analyses), averaged across the two samples (weighted equally).

One caveat is that including those who indicated being only "slightly" liberal or conservative (instead of only those who more strongly identified with either orientation) may have diluted the effect sizes. Nevertheless, these results suggest that the trait content of moral improvement goals is not dramatically different across the political spectrum.

3. Breakdown of Primary Beneficiaries by Trait

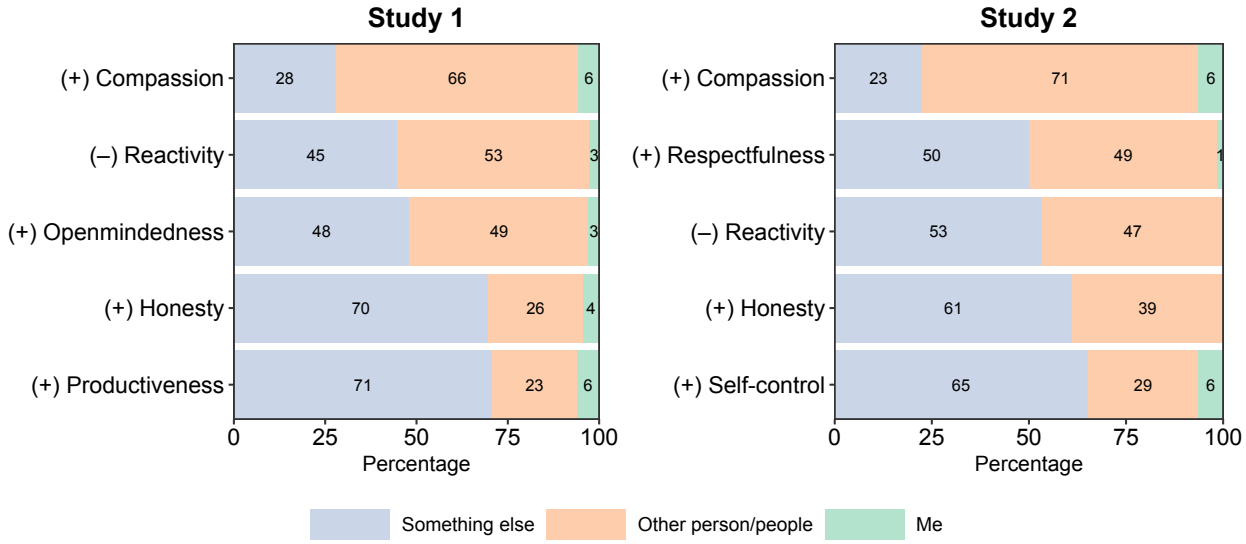
In Study 1, paired-samples *t*-tests showed that participants believed that both the positive (d = 0.17, 95% CI [0.09, 0.25], p < .001) and negative consequences of their moral improvement would be greater for the self than for others (d = 0.45, 95% CI [0.37, 0.52], p < .001). No such differences emerged when examining the perceived well-being impacts for the self as compared with others in Study 2 (d = 0.03, 95% CI [–0.04, 0.11], p = .401). However, when asked for whom their moral improvement would have the most positive consequences (Study 1) or whose lives would be most positively impacted (Study 2), approximately 50% of participants in both studies indicated that they themselves would be the primary beneficiaries of their own moral improvements. This is surprising, given that moral improvements might characteristically be assumed to be undertaken for the benefit of others. To explore whether this counterintuitive finding generalized across different types of moral improvements, we examined the breakdown of the primary beneficiaries across different trait content categories. To ensure sample sizes sufficient to yield reasonably reliable estimates of these proportions, we only examined the breakdowns for the five most frequently mentioned trait content categories in each sample. Sample sizes thus ranged from 90 (More Productive) to 182 (More Compassion) in Study 1, and 63 (More Self-Control) to 239 (More Compassion) in Study 2.

As shown in Figure S2, the percentage of participants who believed that they would be the primary beneficiary of the improvement depended somewhat on the trait content of the moral goal. People were fairly evenly split on whether they or other people would benefit the most if they were to become less reactive (both studies), more open-minded (Study 1), or more respectful (Study 2). Descriptively, a slightly higher percentage of people believed that they (as opposed to others) would benefit the most if they were to become more honest (both studies),

productive (Study 1), or self-controlled (Study 2). However, there was one clear exception: Approximately two thirds of participants in both studies believed that others (as opposed to themselves) would benefit the most from their becoming more compassionate.

Figure S2

Beliefs About Who Would Be Most Positively Impacted by Changing the Five Most Frequently-Mentioned Traits in Studies 1 and 2 (Respectively).



Note. Goals are ordered from the lowest to highest percentage of participants who believed that they would be the primary beneficiaries of that moral improvement.

Supplemental References

Duckitt, J., & Sibley, C. G. (2010). Personality, ideology, prejudice, and politics: A dual-process motivational model. *Journal of Personality*, *78*(6), 1861–1894. https://doi.org/10.1111/j.1467-6494.2010.00672.x

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. In *Advances in experimental social psychology* (Vol. 47, pp. 55-130). Academic Press. https://doi.org/10.1016/B978-0-12-407236-7.00002-4

Lee, K., & Ashton, M. C. (2018). Psychometric properties of the HEXACO-100. *Assessment*, *25*(5), 543–556. https://doi.org/10.1177/1073191116659134

Peterson, C., & Seligman, M. (2004). *Character strengths and virtues: A Handbook and classification*. American Psychological Association.

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