Revisiting the link between true-self and morality: Replication and extension Registered Report of Newman, Bloom, and Knobe (2014) Studies 1 and 2

Shuk Ching (Janet) Lee  
ORCID: 0000-0001-6147-0020  
University of Hong Kong  
[u3591224@connect.hku.hk](mailto:u3591224@connect.hku.hk) / [janetulsc@gmail.com](mailto:janetulsc@gmail.com)

‏‏‏^Gilad Feldman  
ORCID: 0000-0003-2812-6599  
University of Hong Kong  
[gfeldman@hku.hk](mailto:gfeldman@hku.hk) / [giladfel@gmail.com](mailto:giladfel@gmail.com)

^Corresponding author

## 

## Author bios:

Shuk Ching (Janet) Lee was a Master thesis student at the University of Hong Kong during the academic year 2021-2.

Gilad Feldman is an assistant professor with the University of Hong Kong psychology department. His research focuses on judgment and decision-making.

## Declaration of Conflict of Interest:

The author(s) declared no potential conflicts of interests with respect to the authorship and/orpublication of this article.

## Financial disclosure/funding:

This project has been supported by the Teaching Development Grant from the University of Hong Kong awarded to Gilad Feldman.

## Authorship declaration:

Shuk Ching (Janet) Lee conducted the project as part of her MSc thesis in psychology.

Gilad was the thesis advisor and guided Janet. Gilad supervised each step in the project, conducted the pre-registrations, ran data collection, and edited the manuscript for submission.

## Corresponding author

Gilad Feldman, Department of Psychology, University of Hong Kong, Hong Kong SAR; [gfeldman@hku.hk](mailto:gfeldman@hku.hk) ; 0000-0003-2812-6599

## Rights:

CC BY or equivalent license is applied to the AAM arising from this submission. ([clarification](https://bit.ly/rrs-primer))

## Contributor Roles Taxonomy

|  |  |  |
| --- | --- | --- |
| **Role** | **Shuk Ching (Janet) Lee** | **Gilad Feldman** |
| Conceptualization | X | X |
| Pre-registration | X | X |
| Data curation |  | X |
| Formal analysis | X |  |
| Funding acquisition |  | X |
| Investigation | X |  |
| Pre-registration peer review / verification |  | X |
| Data analysis peer review / verification |  | X |
| Methodology | X |  |
| Project administration |  | X |
| Resources |  | X |
| Software | X |  |
| Supervision |  | X |
| Validation |  | X |
| Visualization | X |  |
| Writing-original draft | X |  |
| Writing-review and editing |  | X |

# Abstract

Newman et al. (2014) demonstrated that behaviors that are more aligned with moral values are perceived as more strongly reflecting a person’s “true-self”, suggesting that morality plays an important role in how people perceive others’ essential self. In this Registered Report, we conducted a close replication of Newman et al. (2014)’s Studies 1 and 2 with an online US American sample recruited from Amazon Mechanical Turk using CloudResearch (N = 803). We found support for Study 1’s findings that morally positive changes in others are perceived as more reflective of true-self than morally negative changes, in both the forced-choice (original: η²p = .39, 95% CI [.25, .51]; replication: η²p = .20, 95% CI [.16, .23]) and the continuous scale (original: η²p = .33, 95% CI [.19, .45]; replication: η²p = .22, 95% CI [.15, .25]) measures. We found support for Study 2’s findings that changes more aligned with observers’ political moral views are perceived as more reflective of true-self (original: η²p = .04, 95% CI [.00, .11]; replication: η²p = .35, 95% CI [.29, .41]). Extending the replication, we examined associations between true-self attributions and perceived social norms and found that social norms were positively associated with true self attributions (Study1: most rs ranged from .07 to .21; Study 2: rs = .10 to .30). Materials, data, and analysis code are available on <https://doi.org/10.17605/OSF.IO/9FVTQ>

*Keywords:* True self; social psychology; registered report; replication; morality; moral judgments; social norms; essential self

# Revisiting the link between true-self and morality: Replication and extension Registered Report of Newman, Bloom, and Knobe (2014) Studies 1 and 2

## Background

True self is a mental concept that reflects the deepest and most authentic part of a person’s identity, and people tend to evaluate their true selves as positive. Newman et al. (2014) proposed that attributions of true self in others follow a similar pattern. They demonstrated that morally positive changes in others are perceived as more reflective of their true selves, and that political views guide what people view as morally positive or negative, and so moderate the effect.

In this Registered Report, we conducted a close replication and extension of Newman et al. (2014) with the following goals. Our first goal was to replicate the associations found between morality and true self attributions. Our second goal was to add extensions examining: 1) true-self attributions associations with perceived social norms, and 2) lay-beliefs regarding true-self being inherently good or bad, and comparing these for self versus others.

We begin by introducing the literature on the true self and the chosen article for the replication - Newman et al. (2014). We then review the target article and summarize their hypotheses and findings, and then finally present our adjusted design and suggested extensions.

## True self

True self is defined as the most essential and authentic part of the person’s personality (De Freitas et al., 2017; Schlegel et al., 2011; Strohminger et al., 2017), whereas surface self refers to the more superficial aspects of the self in a person (Newman et al., 2015; Johnson et al., 2004).

There has been increasing interest in the concept of true-self in both the social psychology and experimental philosophy domains (Schlegel et al., 2011; Newman & Knobe, 2019). One common direction of research has been examining associations between true-self and well-being, such as that the subjective feelings of knowing oneself are associated with increased self-esteem and in meaning of life (Schlegel et al., 2009). Overall, the idea of true self seems to be linked with positive aspects for the self, and there is a general tendency for people to evaluate the true-self as positive and moral. Morality is perceived as an essential part of true self (Strohminger & Nichols, 2014; recent replication by Wong & Feldman, 2019), true self attributions are influenced by moral judgments (Kumar, 2016; Strohminger et al., 2017), and people tend to perceive their true selves as morally good (De Freitas et al., 2018; Heiphetz et al., 2017; Strohminger and Nichols, 2014), across ages and cultures (Heiphetz, 2019; De Freitas et al., 2018). This link is helpful in offering some explanations to documented asymmetries in moral judgments (Newman et al., 2015) and is possibly rooted in psychological essentialism (Newman & Knobe, 2019; Strohminger et al., 2017).

## Choice of study for replication: Newman et al. (2014)

We conducted an independent well-powered close replication of Newman et al. (2014)’s Studies 1 and 2, following on the growing recognition of the importance of reproducibility and replicability in psychological science (Nosek et al., 2022; Zwaan et al., 2018). We chose the article by Newman et al. (2014) based on several factors: its academic impact, the potential in methodological improvements and adjustments, and the suitability of its design for adding extensions that would help gain additional valuable insights.

De Freitas et al. (2018) conducted a conceptual replication which seems the closest to Newman et al. (2014), building on their design, examining associations with misanthropy and culture, and reporting a consistent tendency to view the true self as morally good. A recent conceptual replication by Lefebvre and Krettenauer (2020) used a similar design to the target’s Study 1 and concluded that across age groups people do tend to view the true self as moral. We considered these as evidence in support of the phenomenon, yet saw the potential in stronger evidence with a well-powered direct pre-registered replication to try and obtain more precise estimates of the effect size. The reported effects in their Study 1 were very large and likely over-estimated, and as far as we know their Study 2 examining political views as a moderator has not received as much attention with similar conceptual replication attempts.

The target article has had an impact on scholarly research in social psychology, philosophy, judgment and decision-making, and cognitive science (Newman et al., 2015; Kumar, 2016; Strohminger et al., 2017). At the time of writing (April, 2025), there were 356 Google Scholar citations and some important follow-up theoretical and empirical articles, such as Strohminger and Nichols’s (2014) work on the essential moral self, recently successfully replicated by Wong and Feldman (2019).

## Hypotheses and findings in target article

The article by Newman et al. (2014) consisted of three experiments, and we focused our replication on Studies 1 and 2. We chose these studies given that these were the baseline demonstration and more simplified in their design, and given that Study 3 involved aspects of religion, a topic that is considered more sensitive and fast changing in the U.S. American population.

We combined the two studies into a singular data collection, displayed in random order, and made slight adjustments and added extensions to both studies. This design allowed us to both test the designs of the original studies, and to then run further tests in comparing the effects of the different studies with the potential of additional insights. We successfully employed similar designs in previous replications by our team (e.g., Chan & Feldman, 2025; Ding & Feldman, 2025; Wong & Feldman, 2025).

Their Studies 1 and 2 tested two main hypotheses, summarized in Table 1. In their Study 1 the authors hypothesized and demonstrated that others’ morally positive change was more likely than others’ morally negative change to be associated with the true self. In Study 2, the authors predicted and demonstrated that participants’ own moral values determined true self attributions such that changes aligned with political views were more likely to be perceived as reflections of true self. The authors argued that a person's morality is dependent on one’s own views and values (e.g., De Freitas et al., 2017; Graham et al., 2009), which in turn shapes their evaluations of what reflects true-self.

We summarized the findings in the target article in Table 2.

###### Table 1 *Replication and extension: Hypotheses*

|  |  |  |
| --- | --- | --- |
| **Study** | **Hypotheses** | **Description of hypothesis** |
| 1 | 1 (replication) | A morally positive change is perceived as more reflective of true self than a morally negative change or a morally neutral change. |
| 2 | 2 (replication) | Political views moderate the effect, such that change more aligned with liberal values is rated as more reflective of true self by liberals than conservatives, whereas change more aligned with conservative values is rated as more reflective of true self by conservatives than liberals. |
| 1-2 | 3a (extension as exploratory) | Competing hypothesis: Perceived social norms are positively associated with true self and moral attributions. |
|  | 3b (extension as exploratory) | Competing hypothesis: Perceived social norms are negatively associated with true self and moral attributions. |

###### Table 2 *Newman et al. (2014): Summary of findings*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Study | Factors | Reported statistics | | | eta sq | CIL | CIH |
|  |  | *F* | *df* | *p* |  |  |  |
| 1 | Main effect positive-negative on true-self (forced-choice) | 39.92 | 2,127 | < .001 | .39 | .25 | .51 |
| 1 | Main effect positive-negative on true-self (continuous) | 31.01 | 2,127 | < .001 | .33 | .19 | .45 |
| 2 | Interaction between political orientation and conservative-liberal on true-self evaluations (continuous) | 8.44 | 1,199 | = .004 | .04 | .00 | .11 |

*Note*. CIL = lower bounds CIs. CIH = higher bounds CIs.

## Extensions

### Study 1: Morality valence manipulation check

In the target article, the valence of the moral change was assumed yet never directly tested, and so it is possible that some participants perceived items classified under “morally good“ as neutral or even morally bad and “morally bad” items as neutral or even morally good. Furthermore, the target article assumed a clear dichotomy between positive and negative, which greatly simplifies the moral complexity of the items, and limits analyses that consider positive-negative as a continuous scale. We therefore added a morality valence continuous measure as a manipulation check to assess whether participants truly perceive the moral valence of the changes described in the vignettes in the way the experimenters intended, and to allow for testing of associations with other continuous variables.

### Study 1: Continuous true-self and surface-self measures

The target article forced answers using a dichotomy of true self versus surface self. We added continuous measures of true and surface selves to try and gain a clearer more comprehensive understanding of the effect and the distinction between the two.

### Study 2: Vignette political views attribution manipulation check

We added a political views attribution measure as a manipulation check to assess whether participants truly perceive the political view affiliation of the changes described in the Study 2 vignettes in the same way the experimenters intended.

### Study 2: Capturing diverse political orientations

The target article forced a dichotomy of being either liberal or conservative, and by doing so may have failed to capture more complex political categories, possibly resulting in those who do not think of themselves as being conservative or liberal to identify themselves as belonging to one of the two groups. We expected political orientations to be more diverse than the dichotomy used by the target article and therefore expanded the political views options to also allow participants to indicate if they are “independent” or “other”, to try and better capture those who do not self-identify as Conservatives or Liberals. We thought that this adjustment would likely reduce noise and provide for a more accurate test of the hypotheses.

### Study 2: Continuous political orientation measure

Political orientations can be more complex than a simple dichotomy contrasting liberals and conservatives, and we therefore supplemented the categorical political orientations measure with a continuous measure between liberal and conservative, allowing for the midpoint option of being politically “neutral”. We thought that this adjustment has the potential of better capturing complex political views and therefore to more accurately estimate associations between political orientations and attributions.

### Studies 1 and 2: Perceived social norms (exploratory)

We aimed to extend the replication study by examining associations between perceived social norms, true self attributions, and morality. The target article’s reference to morality shifted between examining an absolute positive-negative dichotomy in Study 1 where bad was defined and categorized by the experimenters, to examining individuals’ own moral values in Study 2.

Given the hypothesized link between morality and perceptions of true-self, there are two research questions in respect to social norms. The first is regarding whether one’s morality is aligned with perceived social norms, which may bridge between the different perspectives of morality captured in Study 1 (absolute) versus Study 2 (relativistic). The second is regarding whether perceived social norms are associated with perceived true-self: Is true-self aligned with perceived social norms? True self may be perceived stronger when one follows social norms and social construal of morality, yet it is also possible that true self is perceived stronger when one is perceived as choosing to deviate from social norms and therefore expressing a more free and authentic self separate from others. The link proposed in the target article between morality and true-self implies that adhering to social moral norms and values is associated with stronger perceptions of true authentic self. If that holds true then true-self is seen more in regards to and in alignment with others rather than as differentiating and separate from others. This links with an interesting debate in experimental philosophy and social psychology regarding the purpose of free will (Feldman, 2017; Feldman & Chandrashekar, 2018) with two competing views with one viewing free will as meant for “following rules” in overcoming oneself in order to coexist with others in society, and the second viewing free as meant for allowing for pursuit of one's own wants and needs.

We therefore planned to run an exploratory extension examining associations of morality and true-self perceptions with perceived social norms.

### Studies 1 and 2: Intuitive true self belief (exploratory)

The target article conducted an indirect test whether people perceive true-self to be more aligned with morality and good and bad by asking participants to indicate their perceptions regarding described changes in a person’s character. The implicit nature of the target article’s design introduces several challenges. When evaluating true self by evaluating changes in character, participants might be affected by a variety of factors, such as the feasibility and likelihood of such a change, which may conflict with perceptions of morality which are often considered as an essential stable and durable part of the self (Strohminger & Nichols, 2014).

We therefore added an exploratory extension to supplement the indirect test with an explicit continuous measure directly asking participants about their generalized lay-beliefs regarding the true nature of the self as being good or bad. Using this extension we can examine the alignment between the target article’s implicit test and our more explicit test of the core hypotheses.

Furthermore, we were open to the possibility that laypersons perceive true-self as more complex than a simple dichotomy of good versus bad, as it is possible that people perceive the true self as some mix of both good and bad. We therefore included two separate questions about both good and bad.

In addition, building on a comment by reviewer Dr. Caleb J Reynolds we examined whether perceptions of true-self vary when they are applied to one’s self and when applied to others, with the possibility of finding self-other asymmetries. We therefore examined true-self lay-beliefs both about one’s own true-self and about the average person’s true-self.

## Deviations

We followed the original’s structure of the vignettes, and made slight adjustments to better fit with our target sample and current times. We summarized the deviations in Table 3. First, we neutralized gender and ethnicity in all vignettes, including the opening description and forced-choice measure. The original study began every vignette with the following sentence: “Imagine an individual named \_\_. \_\_ is different from you in almost every way- he has a different occupation and prefers different things than you”. After the amendment, the adjusted unidentified opening description we used was “Imagine someone who is different from you in almost every way- this person has a different occupation and prefers different things than you.” For two specific vignettes like “father” and “boyfriend,” we changed it to “parent” and “romantic partner” respectively. Second, the true self rating in Study 2 was replaced with a 9-point scale used in Study 1 to maintain consistency across the studies.

## 

This Registered Report was written based on the Registered Report template by Feldman (2023).

###### 

###### Table 3 *Replication deviations from the original’s methods and design*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S | Change in… | Original study’s stimuli | Deviation | Justification |
| 1&2 | Scenarios | “Deadbeat father” and “Jerk boyfriend”  “Amir lives in a culture that supports terrorism” | “Father-vignette” replaced “parent”; “boyfriend-vignette” replaced “romantic partner”.  “Amir” was replaced with “someone”. | Addressing possible gender bias and culture bias |
| 1 | Forced choice measure | i) “His “true self” (the deepest, most essential aspect of his being),”  ii) “His “surface self” (the things that he learned from society or others)”,  iii) “None of the above”. | i) “The person’s “true self” (the deepest, most essential aspect of this person’s being)”,  ii) “This person’s “surface self” (the things that this person learned from society or others)”,  iii) “None of the above”. | Addressing possible gender bias |
| 2 | True self rating | A slider bar with “strongly disagree” and “strongly agree” as end points. The corresponding numerical values were 0 and 703. | Replaced with a 9-point scale with “strongly disagree” and “strongly agree” as endpoints. | A more consistent scoring between dependent variables |

# Method

## Reproducibility checks

We calculated effect sizes (ES) and power based on the statistics reported in the target article with the help of a guide by Jané et al. (2024). We ran into minor challenges in our calculations of the effects reported in Study 2, which reported a two-way interaction comparing conservatives to liberals on conservative versus liberal items. Our calculations suggested minor differences from the values reported in the article. For example, our recalculation of the first post-hoc comparisons for conservatives based on the reported t statistic reported resulted in *p* =.007; *d* = .19, 95% CI [0.01, 0.37] rather than the reported *p* = .04, and our recalculations based on the raw descriptives provided (which match with the means in the figure) also seem to result in weaker effects. We consider these rather minor issues, and our current understanding is that these do not change the conclusions of the article. Without access to the raw data and a better understanding of the statistics (correlations between the dependent measures in the mixed-models) it is not possible for us to fully deduce the exact effects.

## Power analysis and sensitivity analyses

We used the R package “pwr”, initially aiming to choose the smallest effect size of the two studies to ensure enough power for all measurements. We provided further information regarding these calculations in the “Effect size calculations and power analysis” subsection in the supplementary materials.

The effect sizes reported in Study 1 were very large (*η*² = 0.39, 95% CI [.25, .51]; *η*² = 0.32, 95% CI [0.19,0.45]), and our power analysis indicated a required sample size of 40 (alpha = .05, power =.95). These were mostly likely overestimated effects. The effect sizes reported in Study 2 were weaker (*η*² = 0.04, 95% CI [.00, 11]; Cohen’s *d* = 0.19 [0.01, 0.37]) and our power analysis indicated that the smallest required sample size was 310.

Given the possibility that the original’s effects are overestimated, even in Study 2, we used the suggested Simonsohn (2015) rule of thumb, even if meant for other designs, and multiplied 310 by 2.5 resulting in 775 participants, aiming for a total sample of 800 participants. A sensitivity analysis using G\*Power (Faul et al., 2007) indicated that a sample of 800 would allow the detection of *f* = 0.06-0.07 (interaction for: groups = 2, measures = 2/3) and *d* = 0.12 for dependent samples t-test contrasts (both 95% power, alpha = 5%, one-tail), effects weaker than any of the supported effects reported in the target article and the standard effects in social psychology for weak effects (Jané et al., 2024).

## Participants

We recruited a total of 803 U.S. American participants from Amazon Mechanical Turk American using CloudResearch MTurk Toolkit (Litman et al., 2017; *Mage*= 43.18, *SD*=12.76; 398 females, 393 males, 13 preferred not to disclose/other). We recruited participants with an approval rate between 95% and 100% and the number of tasks approved between 5,000 and 100,000. We employed the following CloudResearch MTurk Toolkit options which were considered best-practices at the time to ensure high-quality sample: Duplicate IP Block, Suspicious Geocode Block, Verify Worker Country Location, and we recruited only from CloudResearch Approved Participants. We note that 847 subjects began the survey but 44 did not proceed beyond the consent and verifications. We summarized a comparison of study characteristics between the target article and the replication in Table 4.

###### Table 4 *Comparison between the study characteristics between the original study and the replication*

|  |  |  |
| --- | --- | --- |
|  | Newman et al. (2014) | Replication and extension |
| Sample size | Study 1: 130  Study 2: 201 | 803 | |
| Geographic origin | Not specified | U.S. American | |
| Gender | Study 1: 72% female  Study 2: 67% female | 393 males, 398 females, 12 other/did not disclose | |
| Median age (years) | Not specified | 41.0 | |
| Average age (years) | Study 1: 37.0  Study 2: 38.8 | 43.18 | |
| Standard deviation age (years) | Not specified | 12.76 | |
| Age range (years) | Not specified | 21-91 | |
| Medium (location) | Computer (online) | Computer (online) | |
| Compensation | Gift certificates | USD 1.9 | |
| Year | 2014 | 2022 | |

## 

## Design and procedure

We followed the experimental designs by the target article, and summarized the two studies in Tables 5 and 6.

We ran the two studies together in a single data collection, with the display of scenarios and conditions counterbalanced using the randomizer “evenly present” function in Qualtrics. Scenarios were presented in random order and participants were randomly and evenly assigned into the different conditions. This unified design combining replications of several studies into a singular data collection was previously tested successfully in many of the replications and extensions conducted by our team (e.g., Chan & Feldman, 2025; Ding & Feldman, 2025; Wong & Feldman, 2025), and is especially powerful in addressing concerns about the target sample (naivety, attentiveness, etc.) when some studies replicate successfully whereas others do not, as well as in the potential in drawing inferences about the links between the different studies and consistency in participants’ responding to similar paradigms.

Participants first read a consent form and indicated their willingness to participate, and then answered several verification questions. Participants first indicated their consent, with three questions confirming their eligibility, understanding, and agreement with study terms, which they had to answer with a “yes” and the required responses in order to proceed to the study. The three questions also served as attention checks, with a randomized display order of the options - 1) “Are you able to pay close attention to the details provided and carefully answer questions that follow?” (yes/no/not sure), 2) “Do you understand the study outline and are willing to participate in a survey with comprehension checks?” (yes/no/not sure), and 3) “Are you a native English speaker born, raised, and currently located in the US?” (yes/no). Failing any of the three questions meant that the participants did not indicate consent and therefore could not embark on the study. Upon completion of these steps, participants proceeded to begin the survey.

Participants answered the replications of the Studies 1 and 2 in random order. Participants rated true self attributions regarding moralized changes (Study 1: generalized good versus bad changes; Study 2: Moral changes aligned with liberal versus conservative values).

In line with the target article’s design, vignettes in Study 1 were prefixed with a matching of moralized changes in each block so that each block had half negative and half positive changes. We thought this design to be suboptimal compared to a more comprehensive randomization, given that it contrasts specific moral changes against one another, yet decided to follow the target’s design as is.

After completing both experiments, participants rated their political views (used in the replication of Study 2) and their generalized lay-beliefs regarding true self as inherently good and inherently bad (extension). Finally, participants answered funneling questions and provided demographic information, also indicating their level of English understanding of the survey (1 = *Very bad*; 7 = *Very good*), and seriousness in answering the survey (1 = *Not at all*; 5 = *Very much*).

###### Table 5 *Study 1: Summary of experimental design*

|  |  |
| --- | --- |
| **IV1: Block 1 order mix (between-subject)**  Positive change (subset b) + negative change (subset a) + neutral (same)  [Valence within-subject]:   1. Alcoholism-Positive change 2. Boss-Positive change 3. Parent-Positive change 4. Ethnic minorities-Positive change 5. Terrorism-Negative change 6. Business practices-Negative change 7. Romantic partner-Negative change 8. Police officer-Negative change 9. Mac computer-Neutral change 10. Country-Neutral change 11. Cat-Neutral change 12. Football-Neutral change | **IV1: Block 2 order mix (between-subject)**  Positive change (subset a) + negative change (subset b) + neutral (same) [Valence within-subject]:   1. Alcoholism-Negative change 2. Boss-Negative change 3. Parent-Negative change 4. Ethnic minorities-Negative change 5. Terrorism-Positive change 6. Business practices-Positive change 7. Romantic partner-Positive change 8. Police officer-Positive change 9. PC computer-Neutral change 10. City-Neutral change 11. Dog-Neutral change 12. Baseball-Neutral change |
| **True self rating**  DV forced-choice (replication) Please rate what aspect of the person’s personality caused the described change on a choice between:  1) this person’s true self, 2) this person’s surface self, and 3) none of the above.  DV continuous (replication)  Please rate to what extent this person is being true to the deepest, most essential aspects of their being. 0 = *Not at all true to oneself*; 9 = *Very much true to oneself*.  **True self measure**  DV continuous (extension)  Please rate the extent to which the change is a reflection of true self   0 = *Not at all* to 100 = *Completely*  Please rate the extent to which the change is a reflection of surface self   0 = *Not at all* to 100 = *Completely* | |
| **DV: Morality valence** (extension manipulation check)  Do you perceive this person’s change as morally good or morally bad?  -100 = *Very bad*; 0 = *Neither*; 100 = *Very good* | |
| **DV Social norms** (exploratory extension)  Please rate to what extent the described change is in line with the social norms on a scale of -100 to 100 (*very much against social norms* to *very much in line with social norms*). | |
| Preferences towards neutral items in Experiment 1 (replication) [(presented at end of Study 1)] Please indicate your own personal preferences on a 5-point scale with, for example, “strongly prefer dogs” and “strongly prefer cats” as endpoint and “no preference” as the midpoint. | |
| Explicit measures of true-self intuitions (exploratory extension)  [presented at the end after both studies completed)] [shared with Study 2] See “Extensions” under subsection “Studies 1 and 2: Intuitive true self belief” | |

*Note*. IV = Independent variables. DV = dependent variables.

###### Table 6 *Study 2: Summary of experimental design*

|  |
| --- |
| **IV: Condition (within-subject)**  Moral changes in terms of different political orientations  Conservative direction changes (within):   1. Homosexuality change 2. Patriotism change 3. Theism change 4. Monogamy change   Liberal direction changes (within):   1. Global warming change 2. Gender equality change 3. Helping others change 4. Abortion change |
| **DV: True self rating** (replication)  Please rate to what extent at this person’s very essence, there was always something deep within them calling them to\_\_\_ , and then this true self emerged 0 - “*Strongly disagree*”; 9 - “*Strongly agree*” |
| **DV: Political Orientation** (extension manipulation check)  Do you perceive this person’s change as liberal or conservative?  -100 = *Pro-conservative*; 0 = *Neither*; 100 = *Pro-liberal* |
| **Extension DV: Social Norm** (exploratory extension)  Please rate to what extent the described change is in line with the social norms -100 = “*Very much against social norms*” to 100 “*Very much in line with social norms*” |
| Explicit measures of true-self intuitions **(exploratory extension)** [presented at the end after both studies completed)] [shared with Study 1] See “Extensions” under subsection “Studies 1 and 2: Intuitive true self belief” |
| [presented after both studies completed)]  Categorical political measure (adjusted replication) Please choose the one that you feel best represents your political views. “*Conservative*”, “*Liberal*”, “*Independent*”, and “*Other*”. (“Independent” and “other” are adjustments)  Continuous political measure (extension) “Please indicate your political orientation along the conservative-liberal scale” 1 - “*Extremely conservative*”; 4 - “*Center*”; 7 - “*Extremely liberal*”. |

*Note*. IV = Independent variables. DV = dependent variables.

## Measures

### Replication

With the materials sent by the original authors, we were able to reproduce most of the materials in the study. Stimuli for this replication consisted of 12 vignettes from Study 1 and 8 vignettes from Study 2. The opening description for each vignette was: “Imagine someone who is different from you in almost every way - this person has a different occupation and prefers different things than you.”

Each vignette followed the structure that the person used to engage in behavior/belief X and is now involved in behavior/belief Y. In Study 1, changes were framed as good, bad, or neutral. A morally good change was framed such that a behavior/belief changed for the better; a morally bad change was framed such that a behavior/belief changed for the worse. The direction of change was counterbalanced between conditions. Four were changes the authors categorized as morally good, four as morally bad, and four as neutral, and the two exact combinations are provided in Table 5. In Study 2, changes were framed as more aligned with either the conservative or the liberal political views. We followed the original study in classifying the vignettes into binary political ideology: four change vignettes were meant as aligned with conservative views (homosexuality to heterosexuality, unpatriotic to patriotic, atheist to religious, promiscuous to monogamous) and four change vignettes were meant as aligned with liberal views (deny global warming to supporting the environment, sexist to egalitarian, greedy to generous, and vandalizing abortion clinics to not vandalizing abortion clinics).

#### Study 1

##### True self: Forced-choice measure (replication)

Participants indicated their perceptions of whether the change reflected true self with three forced-choice options: a) “true self” (the deepest, most essential aspect of this person’s being), b) “surface self” (the things that this person learned from society or others)”, c) “None of the above” (with a text entry option).

##### True-self: Continuous measure - rating after change (replication)

In Study 1, at the end of each of the 12 vignettes, participants rated whether the person’s final state after the change reflected the person’s true-self (1 = *Not at all*; 9 = *Very much*).

##### Neutral preferences

Preferences on the four neutral items were evaluated on a 5-point scale with, for instance, “*strongly prefer dogs*” and “*strongly prefer cats*” as the endpoints and “*no preference*” as the midpoint.

#### Study 2

##### Continuous true-self rating (replication)

In Study 2, there was a similar question for each of the 8 vignettes with a slight change in describing changes as “the extent to which the change resulted from the emergence of the person’s true self.” (1 = *Strongly disagree*; 9 = *Strongly agree*).

##### Categorical political orientation measure (replication and extension)

We followed the binary political orientation measure in the original study with an extension adjustment of adding two more choices of “other” and “independent.”

### Extensions

#### Study 1: Morality valence manipulation check

In Study 1, we added a manipulation check immediately after the moralized vignettes to assess whether participants assessed the change on a scale from “*Morally bad*” (-100) to “*Neither*” (0) to “*Morally good*” (100).

#### Study 1: Continuous true-self and surface-self measures

Participants responded to what extent to which the change reflects true self and surface self on two separate scales from 0 (*Not at all*) to 100 (*Completely*). Participants answered both scales. This was meant to test both surface and true self separately and as continuous measures.

#### Study 2: Vignette political view manipulation check

In Study 2, we added a manipulation check immediately after the vignettes to examine how participants assessed the changes described in the vignettes: “Do you perceive this person’s change as more pro-liberal or more pro-conservative?” (-100 = *Pro conservative*, 0 = Neither; 100 = *Pro liberal*).

#### Study 2: Continuous political orientation measure

In addition to the categorical political orientation measure, we added a 7-point continuous measure of political orientation (1 = *Extremely conservative*; 4 = *Center*; 7 = *Extremely liberal*).

#### Studies 1 and 2: Perceived social norms (exploratory)

For all vignettes, participants were asked the degree to which the described change of the person was in line with social norms. Participants responded using a -100 to 100 scale with “Very much against social norms” and “Very much in line with social norms” as endpoints.

#### Studies 1 and 2: Intuitive true self beliefs (exploratory)

Participants were asked about their lay-beliefs regarding the nature of true self on a scale of 0 (*Not at all*) to 100 (*Completely*) on two statements : “true self is morally good” and “true self is morally bad”. Participants answered these twice (four items overall), once rating their own true-self (“Please rate your intuitive beliefs regarding your own true self (the deepest and most essential part)” - “my true self is morally good/bad”) and another rating the average person’s true-self (“Please rate your intuitive beliefs regarding the average person’s true self (the deepest and most essential part)” - “average person's true self is morally good/bad”).

## Evaluation criteria for replication findings

We aimed to compare the replication effects with the original effects in the target article using the criteria set by LeBel et al. (2018) (see section “Replication evaluation” in the supplementary materials).

## Replication closeness evaluation

We provided details on the classification of the replication using the criteria by LeBel et al. (2018) in Table 7. We summarize the replication as a "close” replication.

###### Table 7 *Classification of the replication based on LeBel et al. (2018)*

|  |  |  |
| --- | --- | --- |
| **Design facet** | **Replication** | **Details of deviation** |
| Effect/hypothesis | Same |  |
| IV construct | Same |  |
| DV construct | Same |  |
| IV operationalization | Same |  |
| DV operationalization | Same |  |
| Population (e.g. age) | Similar | Data collected using a sample from MTurk using CloudResearch |
| IV stimuli | Different | Neutralized items\* |
| DV stimuli | Similar | Neutralized items and standardized scorings\* |
| Procedural details | Similar | Combined Studies 1 and 2, random order |
| Physical settings | Similar | Online |
| Contextual variables | Similar/Different |  |
| Replication classification | Close replication |  |

*Note.* \*Further details of our deviations can be found in Table 3. IV represents independent variable. DV represents dependent variable.

## 

## Exclusion Criteria

We focused our analyses on the full sample of all participants who completed the study. We had planned to report analyses with exclusions if we failed to find support for the hypotheses (our planned exclusions were: 1) Participants indicating a low proficiency of English (self-report < 5, on a 1-7 scale), and 2) participants who self-report not being serious about filling in the survey (self-report < 4, on a 1-5 scale). Given that we found support for all the hypotheses, we follow the pre-registered Stage 1 plan and do not report additional analyses with exclusions. As an additional exploratory analysis, with our code on the OSF, we also provided the results of our analyses with applying the exclusions, and these had no impact on our findings.

# Results

## Replication

We summarized all descriptive statistics in Tables 8 and 9, and statistical test results in Tables 10 and 11. Plots were created using the JAMOVI (2023) jmv R package.

###### Table 8 *Study 1: Descriptives of true self rating for moralized change (replication + extension)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Conditions |  | Block 1  (*n* = 408)  *M* (*SD*) | Block 2  (*n* = 395)  *M* (*SD*) | **Overall**  **(*N* = 803)**  ***M* (*SD*)** |
| ***Replication:*** Forced-choice measure | | | | |
| Good change |  | 2.47 (1.28) | 2.97 (1.17) | 2.72 (1.25) |
| Bad change |  | 2.16 (1.41) | 2.13 (1.39) | 2.14 (1.40) |
| Neutral change |  | 1.64 (1.24) | 1.27 (1.32) | 1.46 (1.29) |
| ***Replication:*** Continuous true self rating | | | | |
| Good change |  | 6.55 (1.40) | 7.08 (1.29) | 6.81 (1.37) |
| Bad change |  | 5.36 (2.02) | 5.10 (1.90) | 5.23 (1.96) |
| Neutral change |  | 5.83 (1.21) | 5.64 (1.15) | 5.74 (1.18) |
| ***Extension:*** Continuous true self and surface self measures | | | | |
| Good change | True self  Surface self | 64.8 (20.6) 47.3 (23.6) | 72.1 (19.3)  47.5 (23.6) | 68.4 (20.3)  47.4 (23.6) |
| Bad change | True self  Surface self | 56.3 (25.2)  47.7 (24.5) | 53.5 (24.4)  47.5 (23.6) | 54.9 (24.8)  47.6 (24.0) |
| Neutral change | True self  Surface self | 48.0 (21.5) 52.8 (21.5) | 43.2 (22.3)  55.8 (21.9) | 45.7 (22.0)  54.3 (21.8) |

*Note*. *M* indicates mean. *SD* indicates standard deviation. *n*/*N* indicates sample size. There were 4 items for positive, 4 for negative, the calculation for “Forced-choice measure” is the number of items out of the 4 that the participant indicated are a reflection of true-self, therefore range is 0-4.

###### Table 9 *Study 2: Descriptive statistics for true self attribution on changes favoring liberal and conservative values*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Condition | Orientation | *n/N* | *Mean* | *Standard deviation* |
| Overall true self rating  (liberal items) |  | 803 | 5.88 | 1.39 |
| Overall true self rating (conservative items) |  | 803 | 5.51 | 1.58 |
| **Dichotomy political orientation (replication)** | | |  |  |
| True self rating  (liberal items) | Liberal | 414 | 6.11 | 1.39 |
|  | Conservative | 218 | 5.57 | 1.32 |
|  | Independent | 160 | 5.75 | 1.39 |
|  | Other | 11 | 5.39 | 1.24 |
| True self rating  (conservative items) | Liberal | 414 | 4.89 | 1.48 |
|  | Conservative | 218 | 6.63 | 1.33 |
|  | Independent | 160 | 5.64 | 1.26 |
|  | Other | 11 | 5.11 | 1.79 |
| **Continuous political orientation (extension)** | |  |  |  |
| True self rating  (liberal items) | Extremely conservative | 46 | 5.54 | 1.59 |
|  | Very conservative | 96 | 5.57 | 1.30 |
|  | somewhat conservative | 101 | 5.58 | 1.23 |
|  | Center | 122 | 5.81 | 1.36 |
|  | Somewhat liberal | 145 | 6.02 | 1.32 |
|  | Very liberal | 190 | 6.05 | 1.38 |
|  | Extremely liberal | 103 | 6.22 | 1.58 |
| True self rating  (conservative items) | Extremely conservative | 46 | 7.03 | 1.21 |
|  | Very conservative | 96 | 6.81 | 1.22 |
|  | somewhat conservative | 101 | 6.09 | 1.31 |
|  | Center | 122 | 5.65 | 1.24 |
|  | Somewhat liberal | 145 | 5.30 | 1.32 |
|  | Very liberal | 190 | 4.89 | 1.51 |
|  | Extremely liberal | 103 | 4.34 | 1.53 |

*Note.* Mean and standard deviation refers to the descriptive statistics of true self rating on different political vignettes using categorical scale and continuous scale in Study 2.

###### Table 10 *Studies 1 and 2: Summary of statistical tests*

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **S** | **Factor** | **Projects** | ***F*** | ***df*** | ***p*** | ***η²p*** | **CIL** | **CIH** | **Interpretation** | |
| Replication | | | | | | | | | | |
| 1 | Main effect positive-negative on true-self (forced-choice) | Original | 39.92 | 2,127 | < .001 | .39 | .25 | .51 | Signal, inconsistent, smaller |  |
|  | Replication | 199.6 | 2,1602 | < .001 | .20 | .16 | .23 |  |
| 1 | Main effect positive-negative on true-self (continuous rating) | Original | 31.01 | 2,127 | < .001 | .33 | .19 | .45 | Signal, inconsistent, smaller |  |
|  | Replication | 223.7 | 2,1602 | < .001 | .22 | .15 | .25 |  |
| 2 | Interaction between dichotomy political orientation (liberal and conservative) and item types (liberal and conservative) on continuous true self rating | Original | 8.44 | 1,199 | = .004 | .04 | .00 | .11 | Signal, inconsistent, larger |  |
|  | Replication | 340.93 | 1,630 | < .001 | .35 | .29 | .41 |  |
| Extension | | | | | | | | | |  |
| 1 | Main effect positive-negative on true self (continuous true self and surface self measure) |  | 240.5 | 2,1602 | < .001 | .23 | .19 | .27 | Signal |  |
| 1 | Main effect positive-negative on surface self  (continuous true self and surface self measure) |  | 36.93 | 2,1602 | < .001 | .04 | .03 | .06 | Signal |  |
| 2 | Interaction between continuous political orientation and item types (liberal and conservative) on continuous true self rating |  | 260.9 | 1,801 | < .001 | .34 | .28 | .39 | Signal |  |

*Note*. The interpretation of the replication outcomes was based on LeBel et al. (2019) (see section “Additional tables and figures” in the supplementary materials).

###### Table 11 *Studies 1 and 2: Post-hoc tests effect size*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S | Post-Hoc tests | Original  Cohen’s *d*, 95% CI | Replication Cohen’s *d*, 95%CI | Interpretation |
|  | Replication | | |  |
| 1 | Independent sample t-test - forced choice items  (good change vs. bad change) | .53 [.34, .71] | .62 [.47, .76] | Signal, consistent |
| 1 | Independent sample t-test- true self rating  (good change vs. bad change) | .56 [.38, .75] | 1.24 [1.08, 1.39] | Signal, inconsistent, larger |
| 2 | Dependent sample t-test - Liberal participants  (conservative items vs. liberal items) | .19 [.01, .37] | .83 [.72, .94] | Signal, inconsistent, larger |
| 2 | Dependent sample t-test - Conservative participants (conservative items vs. liberal items) | .31 [.09, .54] | .72 [.57, .87] | Signal, inconsistent, larger |
|  | Extension | | |  |
| 1 | Independent sample t-test -continuous true self measure (good vs. bad) | N/A | .87, [.72, 1.01] | Signal |
| 1 | Independent sample t-test -continuous surface self measure (good vs. bad) | N/A | -.02, [-.16, .12] | No signal |

### 

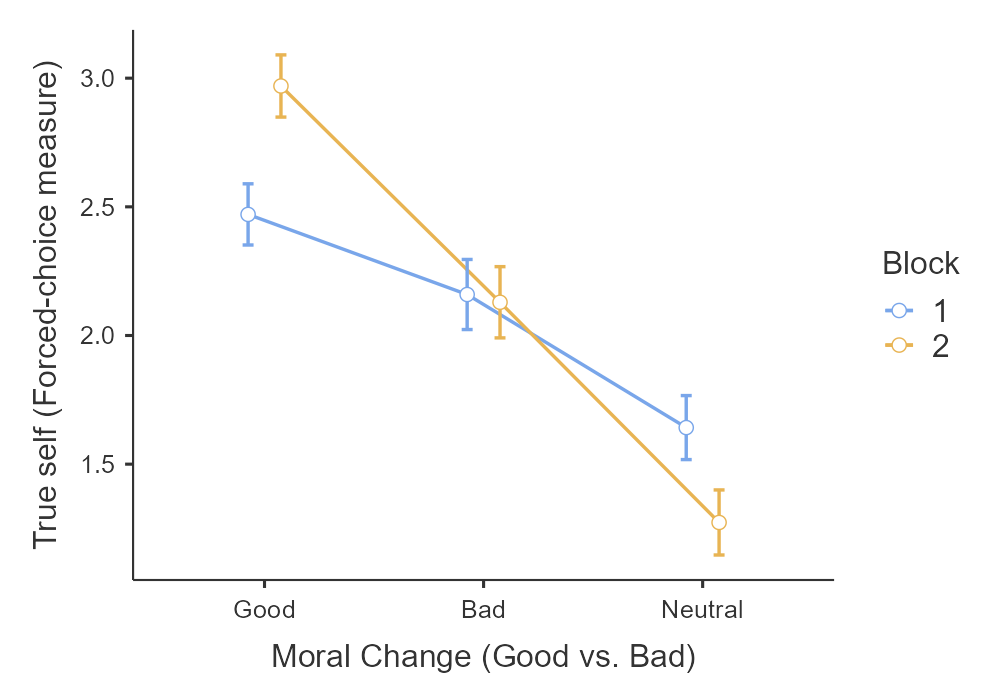
### Study 1: True versus surface self: forced-choice item (replication)

We conducted a 3 (moral valence vignettes: good, bad, neutral; within) x 2 (order: block 1 and block 2; between) repeated-measure ANOVA and found support for a main effect of moral valence (*F*(2, 1602) = 199.6, *p* < .001; *η²p* = .20, 95% CI [.16, .23]), with no main effect for block type, yet support for an interaction with effects being stronger in Block 2 (we reported results of the interaction and block type in the “Additional tables and figures” section of supplementary material). Using the aggregate of the two blocks to mirror the original’s analysis, participants were more likely to rate higher true self for good changes (*M* = 2.72, *SD* = 1.25) than for bad changes (*M* = 2.14, *SD* = 1.40; *t*(801) = 8.72, *p* < .001) (Figure 1), and neutral changes (*M =* 1.46, *SD=* 1.29; *t*(801) = 20.49, *p* < .001; compared to bad changes: *t*(801) = 11.07, *p* < .001). We concluded support for Hypothesis 1 that morally good change is more likely to reveal the true self than morally bad or neutral changes.

Following the original study’s analyses, we performed a chi-squared analysis for all the vignettes separately. We summarized our analyses in Table 12. We found support for participants rating the agent’s true self as more reflected in good changes rather than bad changes among almost all the moralized behaviors, except for the item romantic partner which is consistent with the target article.

We conducted binomial tests comparing the frequency of true self and surface self choices within each vignette to a random 50%-50% split. Consistent with the findings above, we found support for the difference in true self rating for all vignettes compared to the surface self rating in each vignette. Across all morally good vignettes, there were more “true self” choices than “surface self” responses. In comparison, in the morally bad vignettes, only four vignettes indicated more “true self” than “surface self” choices , one indicated more “surface self” choices than “true self”, with the other three showing no support for differences from a 50-50 random split. In the neutral vignettes, five out of eight had more “surface self” than “true self” responses. In general, participants were more likely to attribute true self in good changes than bad changes, except romantic partners, though bad changes had more “true self” than “surface self” responses than we initially expected. We provided more details in the “Additional tables and figures” subsection of the supplementary material.

###### Figure 1 *Study 1: Forced-choice measure result for positive-negative vignettes*



*Note*. True self forced-choice measure across moral-valenced vignettes of good, bad and neutral between block 1 and block 2.

###### Table 12 *Study 1: Chi-squared analysis and independent t-test of forced-choice measure across all vignettes*

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Good change** | | **Bad change** | |  | | **Comparing good to bad on true-self** | | | **Comparing good to bad on surface-self** | | |
| **Moralized Items** | **True self** | **Surface self** | **True self** | **Surface self** | ***X*^2** | ***p*** | ***t*** | ***p*** | **Cohen’s d [LCI, HCI]** | ***t*** | ***p*** | **Cohen’s d [LCI, HCI]** |
| Police officer | 311 | 78 | 206 | 196 | 72.0 | < .001 | -8.73 | < .001 | -.62 [-.76, -.47] | 8.85 | < .001 | .63 [.48, .77] |
| Businessman | 307 | 83 | 210 | 192 | 61.3 | < .001 | -8.07 | < .001 | -.57 [-.71, -.43] | 8.08 | < .001 | .57 [.43, .71] |
| Ethnic Minorities | 269 | 130 | 220 | 168 | 9.80 | .007 | 2.98 | .003 | .21 [.07,.35] | -3.14 | .002 | -.22 [-.36, -.08] |
| Alcoholism | 267 | 133 | 139 | 238 | 73.7 | < .001 | 8.98 | < .001 | .63 [.49, .78] | -8.17 | < .001 | -.58 [-.72, -.43] |
| Terrorism | 305 | 88 | 189 | 210 | 81.5 | < .001 | -9.47 | < .001 | -.67 [-.81, -.52] | 8.97 | < .001 | .63 [.49, .78] |
| Parent | 292 | 111 | 249 | 131 | 9.86 | .007 | 2.59 | .010 | .18 [.04,.32] | -1.84 | .066 | -.13 [-.27,.01] |
| Boss | 180 | 224 | 233 | 152 | 23.0 | < .001 | -4.26 | < .001 | -.30 [-.44, -.16] | 4.72 | < .001 | .33 [.19, .47] |
| Romantic Partner | 250 | 136 | 276 | 119 | 2.94 | .230 | 1.30 | .195 | .09 [-.05, .23] | -1.60 | .110 | -.11 [-.25, .03] |
| Neutral Items | Behavior block 1 (left) | | Behavior block 2 (right) | |  | | Comparing behavior left to right on true self | | | Comparing behavior left to right on surface self | | |
| Mac/PC | 71 | 305 | 72 | 284 | 1.24 | .539 | -.31 | .760 | -.02 [-.16,.12] | .92 | .361 | .06 [-.07, .20] |
| Country/City | 246 | 139 | 130 | 243 | 63.9 | < .001 | 8.07 | < .001 | .57 [.43,.71] | -8.09 | < .001 | -.57 [-.71, -.42] |
| Cat/Dog | 230 | 143 | 185 | 176 | 8.10 | .02 | 2.71 | .007 | .19 [.05, .33] | -2.76 | .006 | -.20 [-.33, -.06] |
| Football/  Baseball | 123 | 258 | 116 | 253 | .06 | .969 | .24 | .809 | .02 [-.12, .16] | -.24 | .811 | -.02 [-.16,.12] |

*Note*. *X*^2 compares the proportions of true-surface self rating in good versus bad. “Behavior block 1 (left)” and “Behavior block 2 (right)” refers to sets of neutral items, where block refers to the block of display and right/left refer to which of the pair is displayed. For example, in the Mac/PC pair, Mac = Behavior block 1 (left), PC = Behavior block 2 (right).

### 

### Study 1: Continuous true self after change rating (replication)

Similarly, we conducted a 3 (moral valence: good, bad, neutral; within) x 2 (order: block 1 and block 2; between) mixed-model ANOVA on the continuous true-self after change ratings and found support for a main effect for moral valence (*F*(2,1602) = 223.7, *p* < .001, *η*²*p* = .22, 95% CI [.15, .25]), no support for a main effect for block type effect, and with support for an interaction (we reported results of the interaction effect and block type effect in the “Additional tables and figures” of supplementary material). We found support for differences between true self ratings for good change (*M =* 6.81*, SD =* 1.37) versus bad change (*M =* 5.23*, SD =* 1.96, *t*(801) = 17.50, *p <* .001) (Figure 2) and compared with the neutral change (*M* = 5.74, *SD* = 1.18; *t*(801) = 19.08, *p <* .001). We again found support for Hypothesis 1 that morally good change is more reflective of true self using the continuous measure.

We summarized a series of t-tests comparing the true self rating on the moral-valence behaviors for all vignettes in Table 13. For all moralized vignettes, we consistently found support that participants were more likely to agree that morally good behavior reflected the agent’s true self than the morally bad behavior. We failed to find support for any difference between the neutral vignettes, except the “country/city” item.

###### Figure 2 *Study 1: Continuous true self ratings for positive-negative vignettes*

**

*Note*. Continuous true self rating across moral-valenced vignettes of good, bad and neutral between block 1 and block 2.

###### Table 13 Study 1: Independent t-test comparing positive-negative on continuous true self ratings

|  |  |  |  |
| --- | --- | --- | --- |
| Moralized items | *t* | *p* | Cohen’s *d* [LCI, HCI] |
| Police officer | 13.3 | < .001 | .94 [-1.09, -.79] |
| Businessman | 12.4 | < .001 | .88 [-1.03, -.73] |
| Ethnic Minorities | 8.00 | < .001 | .56 [.42, .71] |
| Alcoholism | 16.3 | < .001 | 1.15 [.99, 1.31] |
| Terrorism | 14.2 | < .001 | 1.01 [-1.16, -.85] |
| Parent | 10.2 | < .001 | .72 [.57, .86] |
| Boss | 2.79 | .005 | .20 [.06, .34] |
| Romantic Partner | 3.96 | < .001 | .28 [-.42, -.14] |
| Neutral items | *t* | *p* | Cohen’s *d* [LCI, HCI] |
| Mac/PC | -1.19 | .233 | -.08 [-.22, .05] |
| Country/City | 6.38 | < .001 | .45 [.31, .59] |
| Cat/Dog | 2.12 | .034 | .15 [.01, .29] |
| Football/  Baseball | -1.51 | .132 | -.11 [-.25, .03] |

*Note*. In the moralized items, true-self ratings for positive change were always higher than true-self ratings for negative change. In the neutral items, the higher true self ratings were for PC, country, cat, and baseball.

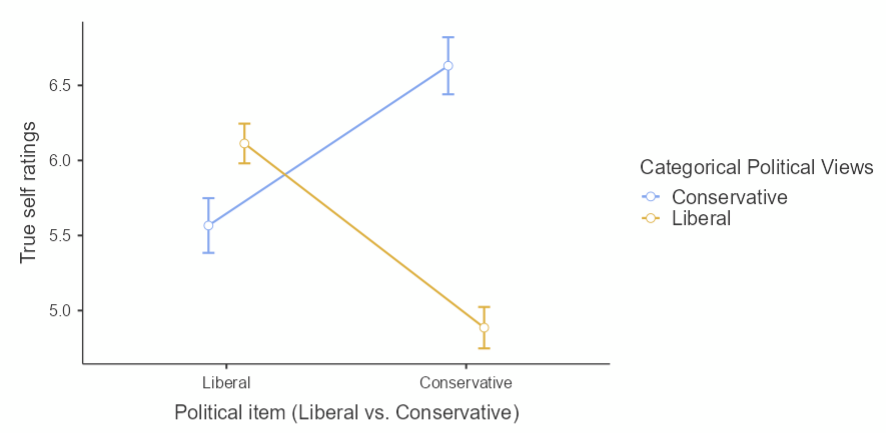
### Study 2: Interaction between political orientation and political item type on true self rating (replication)

We conducted a 2 (political view: liberal and conservative; between) x 2 (item types: liberal and conservative; within) mixed-model ANOVA and found support for an interaction (*F* (1,630) *=* 340.93, *p <* .001, *η*²*p* = .35, 95% CI [.29, .41]), and political view main effect, but no item type main effect (additional details are provided in the subsection under “Additional tables and figures” of supplementary material).

Liberal participants were more likely to agree that the behavior changed resulted from the emergence of a person’s true self for the liberal items (*M* = 6.11, *SD* = 1.39) than for the conservative items (*M* = 4.89, *SD* = 1.48, *t*(630) = 16.84, *p* < .001) (Figure 3). In contrast, conservative participants were more likely to agree that the behavior changed resulted from the emergence of a person’s true self for the conservative items (*M* = 5.57, *SD* = 1.32) than for the liberal items (*M* = 6.63, *SD* = 1.33, *t*(630) = -10.60, *p* < .001). Similar to the original finding, we also found support for a political views main effect that conservative participants (*M* = 6.10, *SD* = 1.09) tended to rate higher overall true-self ratings than liberal participants (*M* = 5.50, *SD* = 1.24, *t* (630) = 6.03, *p* < .001).

Overall, we concluded that we found support for Hypothesis 2 that political views moderated the true self effect, such that changes more aligned with liberal values was rated as more reflective of true self by the liberals than the conservatives, whereas changes more aligned with conservative values was rated as more reflective of true self by the conservatives than the liberals.

###### Figure 3 *Study 2:* *Interaction between political orientation and political item type on true self rating*



*Note*. Continuous true self rating across political items types of liberal and conservative between liberal participants and conservative participants.

## Extensions

### Study 1: Morality valence manipulation check (extension)

We examined whether the participants’ judgment on the different moralized behavior in Study 1 was aligned with the authors’ hypothesized morality in Study 1, a check that was missing in the target article. We expected that: 1) changes categorized as positive would be rated as more positive, and 2) changes categorized as negative would be rated as more negative.

We summarized descriptive and one-sample t-test results in Table 14. We indeed found support for all items being aligned with their assumed valence. We also conducted a series of one-sample t-tests on moralized vignettes and found support for the alignment between participants’ and authors’ moral judgment on the vignettes in Study 1.

###### Table 14 *Study 1: Morality valence manipulation check (one-sample t-test against midpoint 0)*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Item | Good change | | | Bad change | | |
| *M (SD)* | *t* | *p* | *M (SD)* | *t* | *p* |
| Alcoholism | 67.1 (35.9) | 37.7 | < .001 | -51.2 (39.7) | -25.7 | < .001 |
| Boss | 76.0 (34.6) | 44.4 | < .001 | -73.9 (30.8) | -47.7 | < .001 |
| Parent | 83.5 (27.0) | 62.5 | < .001 | -75.0 (33.2) | -45.0 | < .001 |
| Terrorism | 80.4 (29.7) | 53.9 | < .001 | -83.3 (29.5) | -57.0 | < .001 |
| Ethnic Minorities | 82.8 (26.5) | 63.0 | < .001 | -77.7 (34.4) | -44.9 | < .001 |
| Businessman | 79.9 (29.1) | 54.5 | < .001 | -75.8 (31.0) | -49.4 | < .001 |
| Romantic Partner | 79.9 (27.2) | 58.4 | < .001 | -71.8 (35.7) | -40.6 | < .001 |
| Police Officer | 79.7 (31.2) | 50.8 | < .001 | -83.8 (28.0) | -60.5 | < .001 |
|  | Behavior block 1 (left) | | | Behavior block 2 (right) | | |
| Mac/PC | 3.54 (19.1) |  |  | 3.73 (18.0) |  |  |
| Country/City | 13.99 (29.6) |  |  | 4.34 (19.1) |  |  |
| Cat/Dog | 4.60 (17.9) |  |  | 5.08 (17.4) |  |  |
| Football/Baseball | 4.13 (18.8) |  |  | 4.67 (18.7) |  |  |

*Note*. *M* indicates mean. *SD* indicates standard deviation. Scale: -100 to 100, 0 midpoint.  
*n* was either 408 (block 1) or 395 (block 2), depending on the assigned condition for that item.  
“Behavior block 1 (left)” and “Behavior block 2 (right)” refer to sets of neutral items, where block refers to the block of display and right/left refer to which of the pair is displayed. For example, in the Mac/PC pair, Mac = Behavior block 1 (left), PC = Behavior block 2 (right). Valence effects for the positive and negative items were Cohen’s *d* between 1.29 and 3.09.

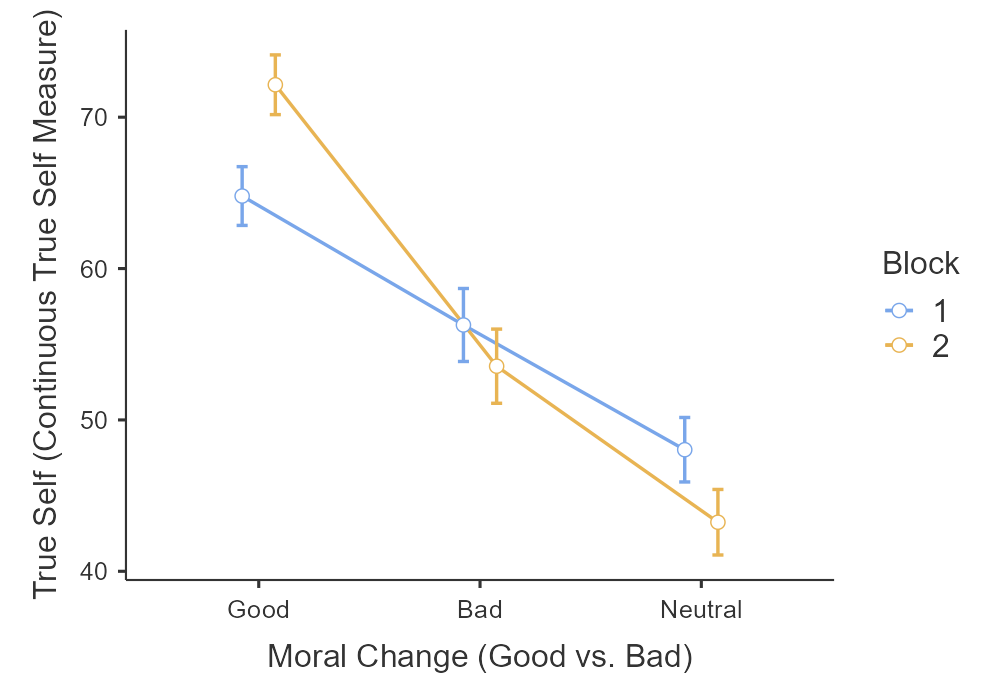
### Study 1: Continuous true-self and surface-self measures (extension)

We added two continuous measures inquiring about what the change reflects, one asked about true-self and the other about surface-self.

Our findings complemented the replication’s continuous true-self after change measure. We first examined the true-self measure and conducted a 3 (moral valence: good, bad, neutral; within) x 2 (order: block 1 and block 2; between) mixed-model ANOVA and found support for a main effect of vignette type (*F*(2, 1602) = 240.5, *p* = .001; *η*²*p* = .23, 95% CI [.19, .27]) and an interaction effect, but not for block type main effect (see subsection “Additional tables and figures” of supplementary material). We again found support for Hypothesis 1 with a larger effect that participants were more likely to attribute true self in morally good changes (*M* = 68.4, *SD* = 20.3) than morally bad change (*M* = 54.9, *SD* = 24.8, *t*(801) = 12.27, *p <* .001) (Figure 4). When compared to neutral change (*M* = 45.7, *SD* = 22.0), true self was more likely to be revealed in morally good change (*t*(801) = 23.56, *p* < .001), and in morally bad change, (*t*(801) = 8.73, *p < .*001).

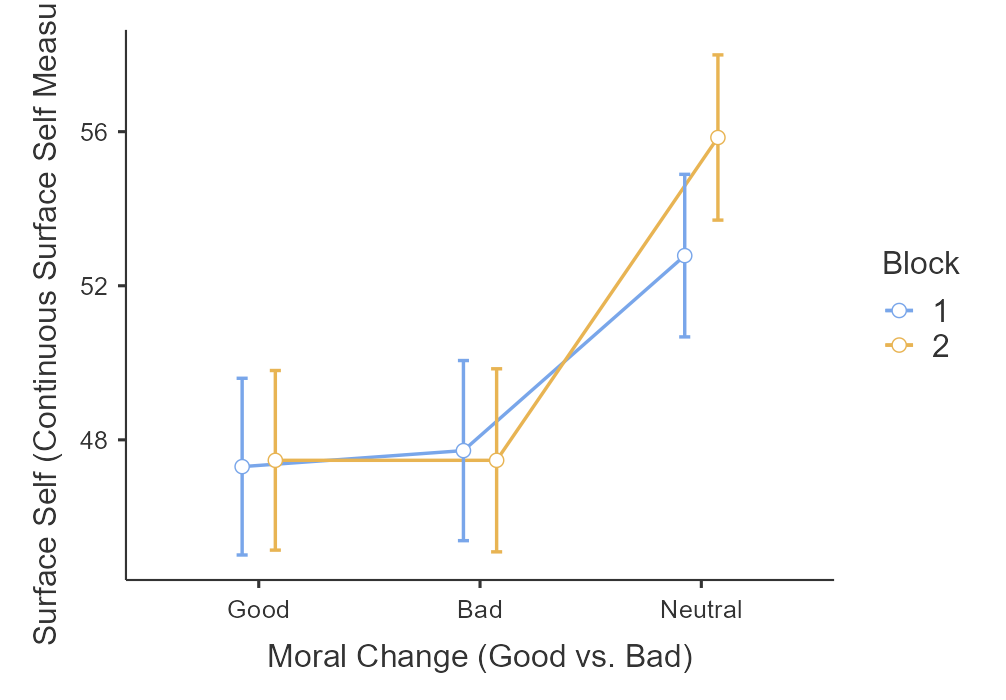
We conducted a similar test on the surface-self measure using a 3 (moral valence: good, bad, neutral; within) x 2 (order: block 1 and block 2; between) mixed-model ANOVA and found support for a main effect of vignette type (*F*(2,1602) = 36.93, *p* < .001; *η*²*p* = .04, 95% CI [.03 .06]; details provided in the “Additional tables and figures” subsection of the supplementary material). Yet, we found no support for differences between surface self ratings for good change (*M =* 47.4*, SD =* 23.6) and bad change (*M* = 47.6, *SD* = 24.0, *t*(801) = -.27, *p* = .962; Figure 5). Neutral change (*M* = 54.3, *SD* = 21.8) was rated as higher surface self compared to morally good (*t(*801) = -7.17, *p* < .001) and morally bad change (*t(*801) = -6.80, *p* < .001). We found no support for the hypothesis that surface self is more reflected in bad change, but rather in neutral changes.

###### Figure 4 *Study 1: Continuous true self measure on moralized changes*



*Note.* Continuous true self measure on good, bad and neutral changes in block 1 and block 2.

###### Figure 5 *Study 1: Continuous surface self measure on moralized changes*



*Note*. Continuous surface self measure on good, bad, and neutral changes in block 1 and 2.

### Study 2: Vignette political view manipulation check (extension)

We examined whether participants' judgments of the political views reflected in the changes in the Study 2 vignettes were aligned with the target article authors’ categorizations. We found support for the target’s categorization that indeed all vignettes aligned with the hypothesized political views. We summarized the descriptive and one-sample t-test results for Study 2 in Table 15.

###### Table 15 *Study2 : Vignette political view manipulation check*

|  |  |  |  |
| --- | --- | --- | --- |
| Item | *M (SD)* | *t* | *p* |
| Conservative change | | | |
| Homosexuality | -35.9 (50.8) | -20.0 | < .001 |
| Patriotism | -41.3 (43.8) | -26.8 | < .001 |
| Theism | -40.1 (45.0) | -25.2 | < .001 |
| Monogamy | -22.1 (39.2) | -16.0 | < .001 |
| Liberal change | | | |
| Global warming | 60.1 (42.2) | 40.4 | < .001 |
| Gender equality | 56.1 (38.7) | 41.1 | < .001 |
| Financial success | 34.0 (45.1) | 21.4 | < .001 |
| Abortion | 50.6 (51.2) | 28.0 | < .001 |

*Note*. M indicates mean. SD indicates standard deviation. *N* = 803. Higher score indicates change is more reflective of liberal rather than conservative political views.

### Study 2: Interaction between continuous political orientation and political item type on true self rating (extension)

We added a continuous political orientation scale as an extension. We conducted a 2 (item types: liberal and conservative; within) repeated ANOVA with a continuous covariate of political orientation measure. We found support the two main effects and the interaction (*F*(1,801) = 409*, p < .*001, *η*²p = .34, 95% CI [.28, .39]) (details provided in subsection “Additional tables and figures” of supplementary material). The findings were consistent with the analysis using the dichotomous political orientation measure. We concluded support for Hypothesis 2 that true self ratings for change depend on alignment with political views.

### Studies 1 and 2: Perceived social norms (exploratory)

First, we tested the associations between perceived social norms and morality in Studies 1 and 2. In Study 1, we found support for positive correlations between social norms and all moralized vignettes (most of *r*s ranged from .40 to .70). In Study 2, for four out of eight politically affiliated vignettes, we found positive correlations between social norms and morality (all *r*s < .20). We found support for Hypothesis 3a that social norms are positively correlated with morality, especially the positive-negative vignettes.

Second, we tested associations between social norms and true self attributions. We found support for the positive correlation between social norms and true self attributions for all the continuous scales (continuous true self rating and continuous true self measure in Study 1, and continuous true self rating in Study 2) (Table 16). In Study 1, true self attributions on the good changes vignettes were positively correlated with norms (all *r*s ranged from .07 to .21). In Study 2, true self rating on both liberal and conservative vignettes were positively correlated with norms (all *r*s ranged from .10 to .30). We found support for the Hypothesis 3a that social norms are positively correlated with true self attributions.

###### Table 16 *Studies 1 and 2: Correlation between perceived social norms and overall true self attributions*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S** | **True self attributions** | **Direction of change** | ***r*** | ***p*** | ***LCH, HCI*** |
| Replication | | | | | |
| 1 | Forced-choice measures | Positive change | -.01 | .825 | -.08, .06 |
|  |  | Negative change | .02 | .527 | -.01, .09 |
| 1 | Continuous true self rating after change | Positive change | .15 | < .001 | .08, .22 |
|  | Negative change | .09 | .015 | .02, .15 |
| 2 | Continuous true self rating | Liberal change | .21 | < .001 | .14, .27 |
|  |  | Conservative change | .16 | < .001 | .09, .22 |
| Extension | | | | | |
| 1 | Continuous true self measure | Positive change | .13 | < .001 | .07, .20 |
|  |  | Negative change | -.03 | .402 | -.10, .04 |
| 1 | Continuous surface self measure | Positive change | .05 | .136 | -.02, .12 |
|  | Negative change | .07 | .039 | .00, .14 |

*Note*. LCH and HCI indicate lower confidence intervals and higher confidence intervals respectively.

### Studies 1 and 2: Intuitive true self beliefs (exploratory)

To supplement the indirect way of assessing the link between true-self and morality, we simply asked participants about their intuitions regarding the true self on the extent to which they thought their and others’ true self is good and is bad. We ran a 2 (self versus others) x 2 (good versus bad) two-way repeated measures ANOVA, and found support for the main effect for valence (*F*(1,802) = 2888.8, *p* < .002, *η*²p = .73), for main effect of target (*F*(1, 802) = 10.0, *p* = .002, *η*²p = .01), and for an interaction (*F*(1, 802) = 518.7, *p* < .001, *η*²p = .39). Ratings of own true self as good (*M* = 82.8, *SD* = 16.5) were far higher than own true self as bad (*M* = 15.0, *SD* = 16.9), with a similar yet weaker effect for others’ (good: *M* = 71.8, *SD* = 17.7; bad: *M* = 27.2, *SD* = 19.6).

We also conducted a correlational analysis exploring the associations between intuitions and true self attributions in Studies 1 and 2. In Table 17 we summarized a comparison of the true self belief of others and true self attributions. We found small to moderate correlations with *r*s ranging from .08 to .28, except the positive true self belief of others on bad change and negative true self belief on conservative change. True self intuitions were positively correlated with true self attributions. In Table 18 we summarized a comparison of the true self belief of one's self and true self attributions. Similarly, we found small to moderate correlations with *r*s ranging from .08 to .28, except the positive true self belief of others on bad change and negative true self belief on conservative change.

###### Table 17 *Studies 1 and 2: Correlation between the true self belief of others with the true self attributions in all vignettes.*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **True self intuitions of others: Positive** | | | **True self intuitions of others: Negative** | | |
| S | Items | *r* | *p* | LCI,HCI | *r* | *p* | LCI,HCI |
| 1 | Forced-choice measure (replication) | | | | | | |
| Good change  Bad change | .12 | < .001 | [.06,.19] | -.10 | .007 | [-.16, -.03] |
| -.09 | .009 | [-.16,.-02] | .12 | < .001 | [.05,.19] |
| 1 | Continuous true self rating (replication) | | | | | | |
| Good change  Bad change | .28 | < .001 | [.22,.34] | -.28 | < .001 | [-.28,-.15] |
| -.13 | < .001 | [-.20,-.06] | .14 | < .011 | [.07,.20] |
| 2 | Continuous true self rating (replication) | | | | | | |
| Liberal change  Conservative change | .19 | < .001 | [.12,.25] | -.09 | .016 | [-.15,-.02] |
| .05 | .161 | [-.02,.12] | .04 | .273 | [-.03,.11] |
| 1 | Continuous true self measure (extension) | | | | | | |
| Good change  Bad change | .25 | < .001 | [.18,.31] | -.16 | < .001 | [-.23,-.10] |
| -.03 | .430 | [-.10,.04] | .08 | .029 | [.01,.15] |

###### Table 18 *Studies 1 and 2: Correlation between one’s own true self belief and the true self attributions in all vignettes*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **True self intuitions on self: Positive** | | | **True self intuitions in self: Negative** | | |
| S | Items | *r* | *p* | LCI,HCI | *r* | *p* | LCI,HCI |
| 1 | Forced-choice measure (replication) | | | | | | |
| Good change | .12 | < .001 | [.06,.19] | -.10 | .007 | [-.16,-.03] |
| Bad change | .09 | .009 | [.16,.02] | .12 | < .001 | [.05,.19] |
| 1 | Continuous true self rating (replication) | | | | | | |
| Good change | .28 | < .001 | [.22,.35] | .22 | < .001 | [-.28,-.15] |
| Bad change | -.13 | < .001 | [-.20,-.06] | -.16 | < .001 | [-.23,-.10] |
| 2 | Continuous true self rating (replication) | | | | | | |
| Liberal change | .24 | < .001 | [.17,.30] | -.13 | < .001 | [-.20,-.06] |
| Conservative change | .17 | < .001 | [.10,.24] | -.07 | .053 | [-.14,.00] |
| 1 | Continuous true self measure (extension) | | | | | | |
| Good change | .25 | < .001 | [.18,.31] | -.16 | < .001 | [-.23,-.10] |
| Bad change | -.03 | .430 | [-.10,.04] | .08 | .029 | [.01,.15] |

## Comparison of replication to original findings

We summarize the comparison of the replication and extension in Tables 10 and 11. We successfully replicated the results for all the chosen studies with Cohen’s *d* condition comparison effects larger than those reported in the original.

# Discussion

In this Registered Report, we conducted a replication of Newman et al. (2014)’s Studies 1 and 2 on true self attributions in value judgment, with added improvements and extensions . We found strong support for all the replication studies and effects.

## Replication of Studies 1 and 2 in Newman et al. (2014)

We were successful in replicating the results by Newman et al. (2014) in support of true self attribution in value judgment. We found that: 1) true self attributions were higher in positive changes than negative or neutral changes; and 2) the effect was moderated by political views. Liberals were more likely to view changes towards liberal views as more reflective of true self than changes towards conservative views, whereas conservatives were more likely to view changes towards conservative views as more reflective of true self than changes towards liberal views. The findings were consistent across multiple measures, using both forced choice and continuous scales. There were some minor inconsistencies, as for example we found bad changes were still regarded more as a reflection of true self than surface self, yet the pattern of lower true-self ratings and higher surface ratings for negative compared to positive held.

Our findings complement work on true-self in support of true self perceived as morally good (Christy et al., 2016; Christy et al., 2017; De Freitas et al., 2017; De Freitas et al., 2018; Fernandez & Schwartz, 2016; Heiphetz et al, 2017; Heiphetz, 2019; Lefebvre & Krettenauer, 2020; Strohminger & Nicholas, 2014). The diagnostic feature of true self is rooted in morality, especially positivity. People seem more likely to agree that deep inside humans are good (Heiphetz et al., 2017). Some studies suggested that positivity is one of the differences between true self and self, such that true self is perceived as good while self can be good or bad (Strohminger et al., 2017). This links with psychological essentialism, which has been used to explain the mechanism behind true self effect (De Freitas et al., 2017, Haslam et al., 2004, Neufeld, 2022). Our tendency to perceive the true self as morally good might be due to the broader tendency to explain things in terms of essences (Christy et al., 2017).

Moral inferences of negative changes were less certain compared to positive changes, possibly due to the destabilizing impact of bad changes made against societal expectations (Siegel et al., 2018; Siedlecki et al., 2013). It is possible that people infer negative changes more cautiously and more diagnostic in updating their views regarding the moral character of an agent (Klein & O’Brien, 2016). Rather than valence alone, the nature of true self seems reflective of what individuals value, and changes going counter to values and perceived social norms seem to also go counter to perceived humans’ true-self. True self might therefore be better described as a dynamic phenomenon taking into account both behavior and environment rather than focused on the person alone.

## Extension: Perceived social norms and intuitive true self belief

In our extension, we found support for the idea that perceived social norms were positively correlated with both morality and true self attributions. The associations with perceived social norms help bridge the theoretical and methodological shift between the two studies in the target article, the absolute morality depicted in Study 1 with clear positive and negative changes, compared with the relativistic morality depending on political views in Study 2.

Our results aligned with work suggesting that true self is more strongly reflected in moral changes than in other conventional or personal changes because of the commonly shared nature of morality (Lefebvre & Krettenauer, 2020). Our extensions were preliminary and exploratory, yet our findings suggest that perceived social norms may play a role in true self attribution, which is somewhat paradoxical, raising the question of how people process the meaning of “true” in “true-self”. If laypersons take “true” to mean “be yourself” then this would seem to mean to be about authenticity and staying true to one’s own direction, honoring the expression of one’s self over and possibly against perceived social norms because it highlights the core part of one’s identity. Our findings suggest otherwise, that the morally good behavior we lead is prone to serve on a pragmatic side of societal consideration instead of a self-enhancing view. The phenomenon of true self as being moral could be interpreted as serving a functional social role to support socially acceptable behaviors in social interaction, promoting good behaviors for human coexistence to control our urges to act in a socially unacceptable way.

To complement the indirect methods assessing true-self attributions regarding valences and moral changes, we added simplified true self intuition measures, and found very consistent results, with very large effects for true-self intuitions. People tend to view the self as being far more good than bad, and they consider themselves as more good and less bad than others, with positive associations between true-self intuitions and true-self attributions. The methodology used in the target article may seem overly complex and long, and some of the participants indicated confusion regarding some of the abrupt changes described in the vignettes, which seems to be a limitation in this commonly used paradigm for testing true self (Strohminger et al., 2017). It is possible that a brief and simple true-self intuitions scale can be used in future research aiming to build on the literature on true-self.

## Limitations and future directions

Our study had several limitations. We focused on the replication, with added extensions that were meant to complement the replication and explore new directions for future research. To ensure the replication was unaffected, we only added dependent measures of social norms to allow us to examine associations with true self attributions, and so our correlational extension findings are only suggestive, and we are unable to infer the causal chain. It seems plausible that true self attributions are affected by perceptions of whether behavior is aligned with social norms, yet it is also possible that norms are adjusted in response to true-self perception goals, such as in adjusting perceived social norms to help maintain a more positive self-image. Future research can build on our findings to do additional experimental work to manipulate social norms and examine how they affect true-self attributions, to examine the inherent conflict in whether being “true” is about being different from or in alignment with others.

Second, we tried to follow the original’s materials as closely as possible, yet we noted that we made several adjustments to the original materials and measures to try and debias from issues like gender and ethnicity. We were successful in our replications, yet it is difficult to estimate how much our changes have impacted the results. Some issues remained unaddressed, also raised in our review process. For example, the forced choice measure of true self in Study 1 might be further improved, as to not force the experimenter’s understanding of true and surface self onto the participants. For example, the original question read “This person’s “surface self” (the things this person learned from society or others)” seems to explicitly imply that surface self reflects learned thoughts or behaviors that are different from the true self. Yet, our findings with other items, seem to suggest that such a description of “surface self” might not always be aligned with how people think of true and surface self, raising the possibility that it is actually true self that is aligned with society and others, not (only) the surface self. Reviewer Dr. Christy A.G suggested that laypersons are unfamiliar with the term “essential” or “non-essential” and that the use of this terminology might further bias responses. Therefore, we suggest future research to carefully rethink the way that true and surface selves are described to get closer to what we aim to study - people’s lay-perceptions. In this specific example, future studies may consider changing the descriptions to “This person’s true self (the deepest, most core aspect of this person’s being)” vs. “This person’s surface self (the shallowest, and more peripheral aspect of this person’s being)” or simply referring to “true self” and “surface self” and letting people infer from that what they will.

In our initial submission we raised concerns regarding the methodological choice in Study 1 to fix the display of items so that each block first displays four positive (/negative) changes together and then four negative (/positive) change vignettes together, followed by four neutral vignettes, which the original authors explained as contrasting certain changes against each other. There were some minor block-order effects that did not seem to impact the overall pattern of results, yet in future research it might be better to randomize the display of the vignettes within each block.

Finally, some of our participants (in the feedback section) and one of our reviewers Dr. Sergio Barbosa expressed concern that the current vignettes did not make any reference to a “mind” behind the described actions. Some participants reflected that there was insufficient information, like the motivation behind changes, to be able to evaluate the true self of the agent for that behavior. The daily value judgment would be different from the fictional change used in the study. It could be more complicated because the judgment on changes might involve other considerations like personal development (Molouki & Bartels, 2017). From a recent review, moral judgment is not centered on the behavior but could be a summary judgment including but not restricted to the mind of the agent such as intention, explanations, and capacities or even the perceived strength of the agent (Hartman et al., 2021). Thus, future research can build on these findings to further explore the role of intent in attributions of true self.

# References

Chan, M., & Feldman, G. (2025). Factors impacting effective altruism: Revisiting heuristics and biases in charity in a replication and extensions Registered Report of Baron and Szymanska (2011). *Royal Society Open Science*.<https://doi.org/10.17605/OSF.IO/BEP78>

Christy, A. G., Seto, E., Schlegel, R. J., Vess, M., & Hicks, J. A. (2016). Straying From the Righteous Path and From Ourselves: The Interplay Between Perceptions of Morality and Self-Knowledge. *Personality and Social Psychology Bulletin*, 42(11), 1538–1550. <https://doi.org/10.1177/0146167216665095>

Christy, A. G., Kim, J., Vess, M., Schlegel, R. & Hicks, J. A. (2017). The Reciprocal Relationship Between Perceptions of Moral Goodness and Knowledge of Others’ True Selves. *Social Psychological and Personality Science*, 8(8), 910–917. <https://doi.org/10.1177/1948550617693061>

De Freitas, J., Cikara, M., Grossmann, I., & Schlegel, R. (2017). Origins of the belief in good   
 true selves. *Trends in Cognitive Sciences*, *21*(9), 634–636.   
 <https://doi.org/10.1016/j.tics.2017.05.009>

De Freitas, J., Sarkissian, H., Newman, G. E., Grossmann, I., De Brigard, F., Luco, A., & Knobe,   
 J. (2017). Consistent belief in a good true self in misanthropes and three interdependent   
 cultures. *Cognitive Science*, *42*, 134–160. <https://doi.org/10.1111/cogs.12505>

De Freitas, J., Tobia, K. P., Newman, G. E., & Knobe, J. (2017). Normative judgments and individual essence. *Cognitive Science*, 41, 382-402. <https://doi.org/10.1111/cogs.12364>

De Freitas, J., Sarkissian, H., Newman, G. Grossmann, I., De Brigard, F., Luco, A., & Knobe, JConsistent belief in a good true self in misanthropes and three interdependent cultures. *Cognitive Science*, *42*, 134-160. <https://doi.org/10.1111/cogs.12505>

Ding, K. , & Feldman, G. (2025). Revisiting Partition Priming in judgment under uncertainty: Replication and extension Registered Report of Fox and Rottenstreich (2003). *Royal Society Open Science*.<https://doi.org/10.17605/OSF.IO/G9CZS>

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>

Feldman, G. (2017). Making sense of agency: Belief in free will as a unique and important construct. *Social and Personality Psychology Compass*, 11(1), e12293.

Feldman, G. (2023). Registered Report Stage 1 manuscript template. <https://doi.org/10.17605/OSF.IO/YQXTP>

Feldman, G., & Chandrashekar, S. P. (2018). Laypersons’ beliefs and intuitions about free will and determinism: New insights linking the social psychology and experimental philosophy paradigms. *Social Psychological and Personality Science*, *9*(5), 539-549. DOI: 10.1177/1948550617713254

Fernandez, D., & Schwartz, B. (2016). Common sense beliefs about the central self, moral   
 character, and the brain. *Frontiers in Psychology*, *6*.   
 <https://doi.org/10.3389/fpsyg.2015.02007>

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029. <https://doi.org/10.1037/a0015141>

<https://doi.org/10.1016/j.copsyc.2021.07.008>

<https://doi.org/10.1177/0146167204271182>

*Journal of Experimental Psychology: General*, *148*(12), 2077. <https://psycnet.apa.org/doi/10.1037/xge0000587>

2017). The role of moral beliefs, memories, and preferences in representations of identity. *Cognitive science*, *41*(3), 744-767. <https://doi.org/10.1016/j.jesp.2018.03.007>

JAMOVI project (2023). *jamovi* (Version 2.4) [Computer Software]. Retrieved from <https://www.jamovi.org>

Jané, M., Xiao, Q., Yeung, S., Ben-Shachar, M. S., Caldwell, A., Cousineau, D., Dunleavy, D. J., Elsherif, M., Johnson, B., Moreau, D., Riesthuis, P., Röseler, L., Steele, J., Vieira, F., Zloteanu, M., & Feldman, G. (2024). Guide to Effect Sizes and Confidence Intervals. <http://dx.doi.org/10.17605/OSF.IO/D8C4G>

Johnson, J. T., Robinson, M. D., & Mitchell, E. B. (2004). Inferences about the authentic self: When do actions say more than mental states?. *Journal of personality and social psychology*, *87*(5), 615. <https://doi.org/10.1037/0022-3514.87.5.615>

Klein, N., & O’Brien, E. (2016). The Tipping Point of Moral Change: When Do Good and Bad   
 Acts Make Good and Bad Actors? *Social Cognition*, 34(2), 149–166.   
 <https://doi.org/10.1521/soco.2016.34.2.149>

Kumar, V. (2016). The empirical identity of moral judgment: Table 1. *The Philosophical   
 Quarterly*, *66*(265), 783–804. <https://doi.org/10.1093/pq/pqw019>

LeBel, E. P., Vanpaemel, W., Cheung, I., & Campbell, L. (2019). A brief guide to evaluate   
 replications. *Meta-Psychology*, *3*. <https://doi.org/10.15626/mp.2018.843>

LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). A unified framework to quantify the credibility of scientific findings. *Advances in Methods and Practices in Psychological Science*, *1*, 389-402.

Lefebvre, J. P., & Krettenauer, T. (2020). Is the true self truly moral? identity intuitions across   
domains of sociomoral reasoning and age. *Journal of Experimental Child Psychology*, *192*, 104769. <https://doi.org/10.1016/j.jecp.2019.104769>

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. Behavior Research Methods, 49(2), 433–442. doi: 10.3758/s13428-016-0727-z

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

Newman, G. E., Bloom, P., & Knobe, J. (2014). Value judgments and the true self. *Personality and Social Psychology Bulletin*, *40*(2), 203–216.

Newman, G. E., De Freitas, J., & Knobe, J. (2015). Beliefs about the true self explain asymmetries based on moral judgment. *Cognitive Science*, *39*(1), 96–125. <https://doi.org/10.1111/cogs.12134>

Newman, G. E., & Knobe, J. (2019). The essence of essentialism. *Mind & Language*, *34*(5), 585–605. <https://doi.org/10.1111/mila.12226>

Neufeld, E. (2022). Psychological essentialism and the structure of concepts. *Philosophy   
 Compass*, *17*(5). <https://doi.org/10.1111/phc3.12823>

Nosek, B.Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., ... & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, 73(1), 719-748. <https://doi.org/10.1146/annurev-psych-020821-114157>

Schlegel, R. J., Hicks, J. A., King, L. A., & Arndt, J. (2011). Feeling like you know who you are:   
Perceived true self-knowledge and meaning in life. *Personality and Social Psychology Bulletin*, *37*(6), 745–756. <https://doi.org/10.1177/0146167211400424>

Schlegel, R. J., Hicks, J. A., Arndt, J., & King, L. A. (2009). Thine own self: true self-concept accessibility and meaning in life. *Journal of personality and social psychology*, *96*(2), 473. <https://psycnet.apa.org/doi/10.1037/a0014060>

Siedlecki, P., Baron, S. G., & Todorov, A. (2013). Diagnostic value underlies asymmetric   
 updating of impressions in the morality and ability domains. *Journal of Neuroscience*,   
 *33*(50), 19406–19415. <https://doi.org/10.1523/jneurosci.2334-13.2013>

Siegel, J., Mathys, C., Rutledge, R., & Crockett, M. (2018). Beliefs about bad people are volatile.   
 *Nature human behaviour*, 2(10): 750–756. DOI: <https://doi.org/10.31234/osf.io/2cqkz>

Simonsohn, USmall Telescopes: Detectability and the Evaluation of Replication Results. Psychological Science, 26(5), 559–569. <https://doi.org/10.1177/0956797614567341>

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

Strohminger, N., & Nichols, S. (2014). The essential moral self. Cognition, 131(1), 159–171. doi: 10.1016/j.cognition.2013.12.005

Wong, C., & Feldman, G. (2019). Essential moral self remains unchallenged: Successful replication of Strohminger and Nichols (2014) with extensions comparing morality to ideology and religion. <https://doi.org/10.17605/OSF.IO/G2ZV6>

Wong, C., & Feldman, G. (2025). Choice Bracketing revisited: Replication and extensions Registered Report of seven experiments reviewed in Read et al. (1999). *Royal Society Open Science*.<https://doi.org/10.17605/OSF.IO/VDQEK>

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, *41*, e120. <https://doi.org/10.1017/S0140525X17001972>