Revisiting the link between true-self and morality: Replication and extension 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 is a Master thesis student at the University of Hong Kong during the academic year 2022.

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:

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

## Authorship declaration:

Shuk Ching (Janet) Lee conducted the replication as part of Thesis in Dissertation in Psychology PSY7308 course.

Gilad was the course instructor for Dissertation in Psychology PSY7308 and led the replication efforts in the course. Gilad supervised each step in the project, conducted the pre-registrations, and ran data collection.

## 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

In the table below, employ CRediT (Contributor Roles Taxonomy) to identify the contribution and roles played by the contributors in the current replication effort. Please refer to <https://www.casrai.org/credit.html> for details and definitions of each of the roles listed below.

|  |  |  |
| --- | --- | --- |
| **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

[IMPORTANT:   
Method and results were written using a randomized dataset produced by Qualtrics to simulate what these sections will look like after data collection. These will be updated following the data collection. For the purpose of the simulation, we wrote things in past tense, but no pre-registration or data collection took place yet.]

People tend to view their own “true self” as generally positive, and as guiding inner moral values. Newman et al. (2014) demonstrated that the true-self link to morality extends also to attributions towards others’ behaviors and changes. We conducted a replication and extensions of Newman et al. (2014)’s Studies 1 and 2, with a US American online Amazon Mechanical Turk sample (*N* = 1000). We found [weak to no] support for Study 1’s findings that morally positive changes in others are perceived as more reflective of true-self than morally negative changes [effect size and CIs]. We found [weak to no] support for Study 2’s findings that changes more aligned with observers’ political moral views are perceived as more reflective of true-self [effect size and CIs]. Extending the replication, we examined associations between true-self attributions and perceived social norms and found that … [effect size and CIs]. Supplementary, materials, raw data and analysis files/code are available here: <https://osf.io/9fvtq/> .

*Keywords:* True self, judgment and decision making, registered replication, moral judgment, social norms

# PCIRR- Study Design Table

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Question | Hypothesis | Analysis plan | Sampling plan | Rationale for the tests | Interpretation given different outcomes | Theory that could be shown wrong by the outcomes |
| What is the relationship between moral change and true self attributions? | True self attributions are higher for morally positive change than for morally negative and morally neutral changes. | Mixed model ANOVA Chi-square test  Binomial test  Independent t-test  Correlation | The current study aimed to recruit 800 participants, well powered enough to detect effects much weaker than the smallest effects in the target. See Power analysis section. | Our replication strategies:  1. We follow the statistical analysis of the original paper.  2. We adopt similar analyses for our extensions. | We examine the replicability of the finding of Newman et al. (2014) based on the criteria used by Lebel et al. (2019). | The link between true-self perceptions and morality. |
| What is the relationship between moral stance and true self attributions? | True self attributions are higher as the changes are more strongly aligned with one’s own moral political views. | Mixed model ANOVA  Independent t-test  Correlation | Moral stance moderates the link between true-self perceptions and morality. |
| What is the relationship between social norms and true self-attributions? | We conducted an exploratory test, competing hypotheses:  1.True-self attributions are positively associated with perceived social norms.  2. True-self attributions are negatively associated with perceived social norms. | Correlation | Exploratory |

# Revisiting the link between true-self and morality: Replication and extension 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 true self attributions of others also follow a similar pattern. They demonstrated that morally positive changes in others are perceived as more reflective of their true selves, and that one’s own moral-political values that guide what people view as morally positive or negative moderate the effect.

We conducted a very close replication of Newman et al. (2014)’s study with the following goals. Our first goal was to replicate the associations between true self attributions, valence, and moral judgment. 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, 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. (2004). 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., 2018; 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 from both the social psychology and experimental philosophy domains (Schlegal 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), true self attributions are influenced by moral judgment (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 (Strohminger et al., 2017; Newman & Knobe, 2019).

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

We chose the article by Newman et al. (2014) based on several factors: its academic impact and the potential for methodological improvements, adjustments, and extensions to gain additional insights.

De Freitas et al. (2018) conducted a conceptual replication which seems the closest to the original’s, building on the design of the target article, 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 well-powered direct pre-registered replications 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. To our knowledge, there are currently no independent direct close pre-registered replications of the article.

The target article has had an impact on scholarly research in social psychology, philosophy, judgment and decision-making, and cognitive science (Strohminger et al., 2017; Newman et al., 2015; Kumar, 2016). At the time of writing, there were 246 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.

We aimed to revisit the classic phenomenon to examine the reproducibility and replicability of the findings with an independent replication. Following the recent growing recognition of the importance of reproducibility and replicability in psychological science (Brandt et al., 2014; Nosek & Errington, 2020; Veer & Giner-Sorolla, 2016; Zwaan et al., 2018), we embarked on a well-powered pre-registered close replication of Newman et al. (2014)’s Studies 1 and 2.

## 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 one specific topic that is more sensitive and fast changing in the American population and with references to religion.

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 in our team (e.g., Adelina & Feldman, 2022; Vonasch et al., 2022; Yeung & Feldman, 2022).

Their Studies 1 and 2 tested two main hypotheses, summarized in Table 1. Given the notion of morally good true self suggested in previous studies, 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. The authors argued that people have a deep-rooted motivation to act morally, and that people’s morality affects perceptions of various psychological states: intentionality (Cushman & Mele, 2008; Leslie et al., 2006), happiness (Phillips et al., 2011), and preferences towards outcomes (Pettit & Knobe, 2009). In Study 2, the authors predicted and demonstrated that participant’s 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., Graham et al., 2009; De Freitaset al., 2017), which in turn shapes their evaluations of what reflects true-self.

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

Table 1

*Summary of the 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) | Perceived social norms are positively associated with true self and moral attributions. |
|  | 3b (extension as exploratory) | Perceived social norms are negatively associated with true self and moral attributions. |

Table 2

*Summary of original findings in Newman et al. (2014)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 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 for CIs. CIH = higher bounds of CIs.

## Extensions

### Study 1: Morality valence manipulation check

We 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 same way the experimenters did, and to test continuous associations. The direction of change was assumed but never directly tested, and so it is possible that some participants perceive some of the described morally good changes as neutral or even morally bad. Furthermore, it assumes a clear dichotomy between positive and negative, which limits analyses that consider positive-negative as a continuous scale.

### 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 view manipulation check

We added a political view measure as a manipulation check to assess whether participants truly perceive the changes described in the vignettes as politically affiliated in the same way the experimenters did.

### 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 other 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. This is likely to reduce noise and provide for a better test of the hypotheses.

### Study 2: Continuous political orientation measure

Political orientations can be more complex than a simple contrast between liberal and conservative, 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”. This has the potential of better capturing and therefore a more accurate estimate of the associations with political orientations.

### 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; Nanakdewa et al., 2022) 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 as 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 is 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 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.

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 |

## Pre-registration and open-science

[*Note: Written in past tense, but is just a reflection of what it would look like after data collection*]

We pre-registered the experiment on the Open Science Framework (OSF) and data collection was launched later that week. Pre-registrations, power analyses, and all materials used in these experiments are available in the supplementary materials. We provided all materials, data, code, and pre-registration on the OSF: <https://osf.io/9fvtq/>. We provided additional open-science details and disclosures in the supplementary materials under “Open Science disclosures” sub-section.

All measures, manipulations, exclusions conducted for this investigation are reported, all studies were pre-registered, and data collection was completed before any analyses.

# Method

[IMPORTANT:   
Method and results were written using a randomized dataset produced by Qualtrics to simulate what these sections will look like after data collection. These will be updated following the data collection. For the purpose of the simulation, we wrote things in past tense, but no pre-registration or data collection took place yet.]

## Reproducibility checks

We calculated effect sizes (ES) and power based on the statistics reported in the target article. 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, which according to our current understanding 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 conducted calculated effects in the target article using 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 are 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. Accounting for possible exclusions and the integrated design, and allowing for the potential of additional analyses, we aimed for a total sample of 800 participants. A sensitivity analysis using 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.

To demonstrate what the results would look like after data collection we simulated a dataset of 1000 participants using Qualtrics, which we will later update with the real data and our sample of ~800.

## Participants

We generated a sample of 1000 dummy participants (*M*age= 50, *SD* = 29.2; 246 females, 257 males, 497 prefers not to say/other.)

We will recruit participants from Amazon Mechanical Turk using the CloudResearch/Turkprime platform (Litman et al., 2017). Based on our extensive experience of running similar replications on MTurk, to ensure high quality data collection, we employed the following CloudResearch options: Duplicate IP Block. Duplicate Geocode Block, Suspicious Geocode Block, Verify Worker Country Location, Enhanced Privacy, CloudResearch Approved Participants, Block Low Quality Participants, etc. We will also employ the [Qualtrics fraud and spam prevention measures](https://www.qualtrics.com/support/survey-platform/survey-module/survey-checker/fraud-detection/): reCAPTCHA, prevent multiple submission, prevent ballot stuffing, bot detection, security scan monitor, relevantID, etc.

Assignment pay is based on the federal wage of 7.25USD/hour, per minute, so for example - 5-8 minutes survey would be paid 1USD per participant. We first pretest survey duration with 30 participants to make sure our time run estimate was accurate and then adjust pay as needed, the data of the 30 participants will not be analyzed separately from the rest of the sample other than to assess survey completion duration and needed pay adjustments. For those pretest participants, if survey duration was longer than expected, they would be paid a bonus as pay adjustment.

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 | 1000 | |
| Geographic origin | Not specific | US American | |
| Gender | Study 1: 72% female  Study 2: 67% female | 257 males, 246 females, 497 other/did not disclose | |
| Median age (years) | Not specific | 50.0 | |
| Average age (years) | Study 1: 37.0  Study 2: 38.8 | 50.0 | |
| Standard deviation age (years) | Not specific | 29.2 | |
| Age range (years) | Not specific | 0.00-100 | |
| Medium (location) | Computer (online) | Computer (online) | |
| Compensation | Gift certificates | Not applicable | |
| 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 will run 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 will be presented in random order and participants will be randomly and evenly assigned into the different conditions.

This method was previously tested successfully in many of the replications and extensions conducted by our team (e.g., Adelina & Feldman, 2022; Vonasch et al., 2022; Yeung & Feldman, 2022), and is especially powerful in addressing concerns about the target sample (naivety, attentiveness, etc.) when some studies replicate successful 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.

## Procedure

[*For review: The Qualtrics survey .QSF file and an exported DOCX file are provided on the OSF folder. A preview link of the Qualtrics survey is provided on:* [*https://hku.au1.qualtrics.com/jfe/preview/SV\_brLyDVuvvTEvGiG?Q\_CHL=preview&Q\_SurveyVersionID=current*](https://hku.au1.qualtrics.com/jfe/preview/SV_brLyDVuvvTEvGiG?Q_CHL=preview&Q_SurveyVersionID=current)]

We reached out to the authors of the target article and are grateful for the materials they provided which were helpful in our reconstruction of the materials. They have also been responsive and supportive in follow-up interactions.

Following consent and qualification questions, participants answered the replications of the Studies 1 and 2 the target article 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*

|  |  |
| --- | --- |
| **Individual differences**  Preferences towards neutral items in Experiment 1 (replication) 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 (extension) See “Extensions” under subsection “Studies 1 and 2: Intuitive true self belief” | |
| **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**  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 (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). | |

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

Table 6   
*Study 2: Summary of experimental design*

|  |
| --- |
| **Individual differences**  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”.  Explicit measures of true-self intuitions (exploratory extension) See “Extensions” under subsection “Studies 1 and 2: Intuitive true self belief” |
| **IV2: 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 |
| **Replication DV: True self rating**  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)**  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*” |

*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, and neutral. A morally good change means a behavior/belief changed from bad to good; a morally bad change means a behavior/belief changed from good to bad. 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 favorable to either conservative or liberal views. We followed the original study in classifying the vignettes into binary political ideology: four change vignettes were in support of conservative views (homosexuality to heterosexuality, unpatriotic to patriotic, atheist to religious, promiscuous to monogamous) and four change vignettes were in support of 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

##### 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 rating current state (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 + extension)

We followed the binary political orientation measure in the original study with an extension adjustment of adding two more choices as “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 were allowed to answer 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 ocial 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).

## 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 from MTurk |
| 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 MTurk |
| 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.

## 

We followed the data analysis strategy as in the original byWe performed correlational analyses to examine the associations between social norms, morality, and true self attributions and between true self beliefs and true self attributions.

## 

We focus on our analyses on the full sample of all participants who completed the study. However, if we fail to find support for the hypotheses we will rerun our analyses with exclusions, and report a comparison of the findings pre and post (see “Pre-exclusions versus post-exclusions” section in the supplementary materials). We will consider these exploratory. Our planned exclusions in case of failure at this stage are: 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).

# Results

[IMPORTANT:   
Method and results were written using a randomized dataset produced by Qualtrics to simulate what these sections will look like after data collection. These will be updated following the data collection. For the purpose of the simulation, we wrote things in past tense, but no pre-registration or data collection took place yet.]

## Replication

We summarized all descriptive statistics in Tables 7 and 8, and planned statistical tests in Table 9.

[We provided a detailed JAMOVI data analysis file on our simulated data in the OSF directory (.OMV and output in .PDF), for a full description of the planned data analyses]

Table 7

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Conditions |  | | Block 1  (*n*=500)  *M* (*SD*) | Block 2  (*n*=500)  *M* (*SD*) | Overall  (*N*=1000)  *M* (*SD*) |
| ***Replication:*** Continuous true self rating | |  |  |  |  |
| Good change |  | | 4.91 (1.33) | 4.94 (1.28) | 4.92 (1.30) |
| Bad change |  | | 4.94 (1.27) | 4.92 (1.35) | 4.93 (1.31) |
| Neutral change |  | | 4.95 (1.34) | 4.92 (1.29) | 4.94 (1.32) |
| ***Extension:*** Continuous true self and surface self measure | | | | | |
| Good change | True self  Surface self | | 49.2 (14.5) 50.0 (14.6) | 50.0 (15.2)  49.9 (15.0) | 49.6 (14.9)  49.9 (14.8) |
| Bad change | True self  Surface self | | 50.1 (14.4)  50.0 (15.0) | 49.6 (14.5)  49.9 (15.0) | 49.9 (14.4) 49.9 (15.0) |
| Neutral change | True self  Surface self | | 49.7 (14.6) 50.8 (14.2) | 50.5 (15.0)  50.3 (14.9) | 50.1 (14.8) 50.6 (14.5) |



*Note*. M indicates mean. SD indicates standard deviation. *n*/*N* indicates sample size.

Table 8

*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) |  |  | 4.93 | 1.28 | |
| Overall true self rating (conservative items) |  |  | 4.99 | 1.25 | |
| **Dichotomy political orientation (replication)** | | |  |  |  |
| True self rating  (liberal items) | Liberal | 234 | 4.97 | 1.26 | |



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Conservative | 251 | 5.00 | 1.25 |
|  | Independent | 259 | 4.91 | 1.33 |
|  | Other | 256 | 4.84 | 1.28 |
| True self rating  (conservative items) | Liberal | 234 | 5.00 | 1.23 |
|  |  | 251 | 4.98 | 1.25 |
|  | Independent | 259 | 4.93 | 1.17 |
|  | Other | 256 | 5.05 | 1.33 |
| **Continuous political orientation (extension)** | |  |  |  |
| True self rating  (liberal items) | Extremely conservative | 124 | 5.01 | 1.32 |
|  | Very conservative | 136 | 4.91 | 1.28 |
|  | somewhat conservative | 136 | 4.90 | 1.27 |
|  | Center | 153 | 4.94 | 1.31 |
|  | Somewhat liberal | 148 | 4.94 | 1.20 |
|  | Very liberal | 152 | 4.88 | 1.31 |
|  | Extremely liberal | 151 | 4.93 | 1.30 |
| True self rating  (conservative items) | Extremely conservative | 124 | 4.95 | 1.17 |
|  | Very conservative | 136 | 5.01 | 1.22 |
|  | somewhat conservative | 136 | 5.06 | 1.23 |
|  | Center | 153 | 4.95 | 1.33 |
|  | Somewhat liberal | 148 | 5.26 | 1.17 |
|  | Very liberal | 152 | 4.84 | 1.23 |
|  | Extremely liberal | 151 | 4.89 | 1.32 |

*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 9

*Studies 1 and 2: Summary of statistical tests*

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Replication*** | | | | | | | | | | | |
| S | | Factor | Projects | *F* | *df* | *p* | eta square | CIL | CIH | Interpretation |  |
| 1 |  | Main effect positive-negative on true-self (forced-choice) | Original | 39.92 | 2,127 | < .001 | .39 | .25 | .51 | N/A |  |
|  | | Replication | 1.01 | 2,1996 | .364 | .00 | .00 | .00 |  |  |
| 1 | | Main effect positive-negative on true-self (continuous rating) | Original | 31.01 | 2,127 | < .001 | .33 | .19 | .45 | N/A |  |
|  | | Replication | .03 | 2,1996 | .969 | .00 | .00 | .00 |  |  |
| 2 | | Interaction between dichotomy political orientation (liberal and conservative) and item types (liberal and conservative) on continuous true self rating | Orignal | 8.44 | 1,199 | = .004 | .04 | .00 | .11 | N/A |  |
|  | | Replication | .09 | 1,483 | .764 | .00 | .00 | .00 |  |  |
| ***Extension*** | | | | | | | | | | |  |
| 1 |  | Main effect positive-negative on true self (continuous true self and surface self measure) | Current | 0.31 | 2,1996 | .737 | .00 | .00 | .00 | N/A |  |
| 1 |  | Main effect positive-negative on surface self  (continuous true self and surface self measure) | Current | .71 | 2,1996 | .491 | .00 | .00 | .00 | N/A |  |
| 2 |  | Interaction between continuous political orientation and item types (liberal and conservative) on continuous true self rating | Current | .03 | 1,998 | .856 | .00 | .00 | .00 | N/A |  |



*Note*. The interpretation of outcomes will be based on LeBel et al. (2019) on actual data.

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

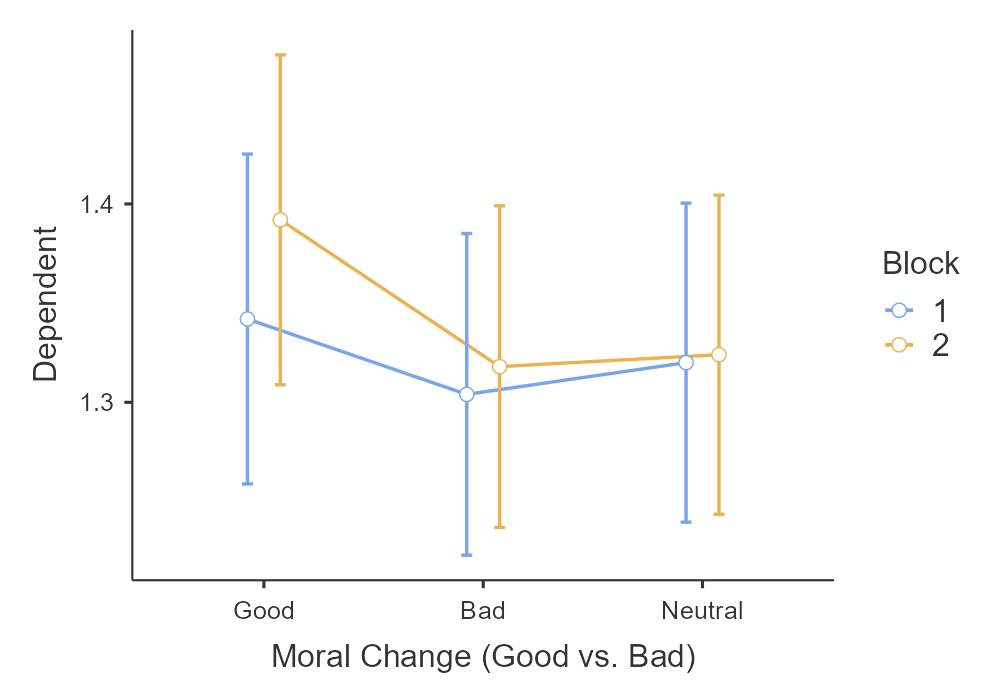
We conducted a 3 (moral valence vignettes: good, bad, neutral; within) x 2 (order: block 1 and block 2; between) repeated-measure ANOVA and found no support for a main effect of valence, *F*(2,1996) = 1.01, *p* = .364; η² = .00; 95% CI [.00, .00], or block type, or their interaction. We found no support for differences in rating good change (*M* = 1.37, *SD* = .95) or bad change (*M* = 1.31, *SD*=.92) as true self (*M*diff = .02, *SE* =.04, *t*(998) = 1.35, *p* = .366) (Figure 1). Similarly, we found no support when compared to neutral change (*M* = 1.32, *SD* = .92; *M*diff = .05, *SE* =.04, *t*(998) = 1.07, *p* = .531). We failed to find support for the hypothesis that morally good change is more likely to reveal the true self than morally bad change.

Following the original study’s analyses, we performed a chi-squared analysis (with the true self, surface self, and others as the three potential choices) for all the vignettes. As shown in Table 10, we found no support for the difference in attributing the true self in morally good and morally bad change among all moralized behaviors.

We conducted binomial tests to examine the frequency of true self and surface self choices within each vignette.We found no support for the difference in true self rating for all vignettes compared to the surface self rating in each vignette. Please find the detail under the subsection of “Additional tables and figures” in supplementary material.

Figure 1

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



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

10

*Study 1: Chi-squared analysis and independent t-test of true self rating 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 | 169 | 155 | 177 | 167 | *1.84* | *.399* | *-.53* | *.595* | -.03[-.16,.09] | *-.81* | *.417* | *-.05[-.18,.07]* |
| Businessman | 162 | 166 | 175 | 164 | *.88* | *.645* | *-.87* | *.385* | -.05[-.18,.07] | *.13* | *.893* | *.00[-.12,.13]* |
| Ethnic Minorities | 179 | 161 | 156 | 173 | *1.83* | *.400* | *1.20* | *.230* | .08[-.05,.20] | *-1.14* | *.253* | *-.07[-.20,.05]* |
| Alcoholism | 167 | 179 | 167 | 161 | *1.95* | *.378* | *.00* | *1.000* | .00[-.12,.12] | *1.20* | *.230* | *.08[-.05,.20]* |
| Terrorism | 171 | 165 | 171 | 161 | *.10* | *.953* | *.00* | *1.000* | .00[-.12,.12] | *.27* | *.788* | *.01[-.11,.14]* |
| Parent | 169 | 172 | 177 | 173 | *.45* | *.799* | *-.53* | *.595* | -.03[-.16,.09] | *-.07* | *.947* | *-.00{-.13,.12]* |
| Boss | 156 | 179 | 154 | 178 | *.04* | *.979* | *.14* | *.891* | .01[[-.12,.13] | *.07* | *.947* | *.00[-.12,.13]* |
| Romantic Partner | 150 | 160 | 173 | 180 | *8.3* | *.016* | *-1.56* | *.120* | -.10[-.22,.03] | *-1.33* | *.182* | -.08[-.21,.04] |
| Neutral Items | Behavior block 1 (left) | | Behavior block 2 (right) | |  | |  | | |  | | |
| Mac/PC | 158 | 176 | 160 | 170 | *.16* | *.921* | *-.136* | *.892* | -.01[-.13,.12] | *.40* | *.690* | *.03[-.10,.15]* |
| Country/City | 163 | 172 | 157 | 174 | *.17* | *.918* | *.41* | *.685* | .03[-.10,.15] | *-.13* | *.894* | *-.01[-.13,.12]* |
| Cat/Dog | 178 | 165 | 172 | 181 | *1.17* | *.557* | *.40* | *.691* | .03[-.10,.15] | *-1.06* | *.288* | *-.07[-.19,.06]* |
| Football/  Baseball | 161 | 186 | 173 | 161 | *2.76* | *.251* | *-.80* | *.422* | -.05[-.18,.07] | *1.66* | *.097* | *.11[-.02,.23]* |

*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).

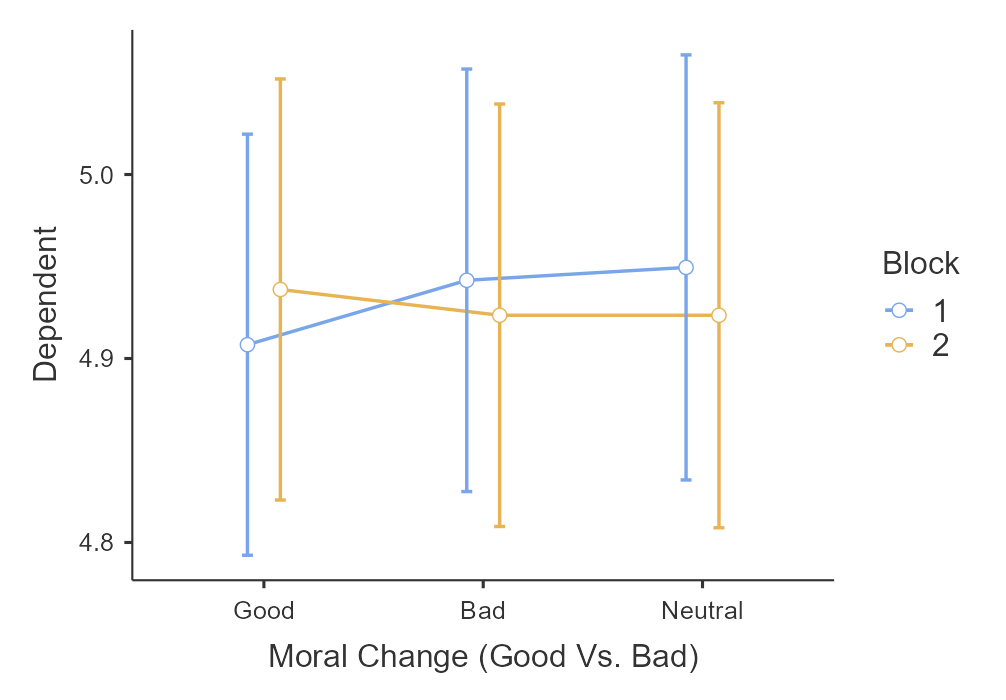
### Study 1: Continuous true self rating (replication)

Similarly, we conducted a 3 (moral valence: good, bad, neutral; within) x 2 (order: block 1 and block 2; between) mixed-model ANOVA and found no support for a main effect of vignette type (*F*(2,1996) = .03, *p* = .969, *η*²= .00, 95%CI [.00, .01]), block type, or interaction. We found no support for differences between true self ratings for good change (*M* = 18.3, *SD* = 5.4) versus bad change (*M* = 18.0, *SD* = 5.28; *M*diff = -.01, *SE*=.06, *t*(1996) = -.18, *p* = .983) (figure 2). We failed to support the hypothesis that morally good is more likely to reveal the true self.

We summarized a series of t-tests comparing the true self rating on the moral-valence behaviors for all vignettes in Table 11. We found no support for differences across all the vignettes.

Figure 2

*Study 1: Continuous true self rating 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 11

*Study 1: Independent t-test comparing positive-negative on continuous true self rating*

|  |  |  |  |
| --- | --- | --- | --- |
| Moralized items | *t* | *p* | Cohen’s d [LCI, HCI] |
| Police officer | .64 | .521 | .04 [-.08,.17] |
| Businessman | .04 | .971 | .00 [-.12,.13] |
| Ethnic Minorities | .22 | .824 | .01 [-.11,.14] |
| Alcoholism | -.11 | .912 | -.01 [-.13,.12] |
| Terrorism | -1.24 | .214 | -.08 [-.20,.05] |
| Parent | -.39 | .695 | -.02 [-.15,.10] |
| Boss | -.11 | .912 | -.01 [-.13,.12] |
| Romantic Partner | .68 | .497 | .04 [-.08,.17] |
| Neutral items |  | | |
| Mac/PC | -.31 | .760 | -.02 [-.14,.12] |
| Country/City | 1.65 | .10 | .10 [-.02, .23] |
| Cat/Dog | -.56 | .577 | -.04 [-.16,.09] |
| Football/  Baseball | -.17 | .863 | -.01 [-.14,.11] |

### Study 1: Neutral items and preferences (replication)

We ran raw correlations and a one-way ANOVA on the neutral items and neutral item preferences. We found no support for a relationship between neural item preference and the neutral item true self ratings.

### 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 no support for an interaction effect (*F*(1,483) = .09, *p* = .764, *η*²= .00, 95% CI [.00, .00]). For the conservative items, we found no support for differences in true self ratings between conservatives (*M* = 4.98, *SE=.08*) and liberals (*M* = 5.00, *SE=.08*; *Mdiff*= -.02, *t*(483) = -.17, *p* = .998). Similarly, for the liberal items, we found no support for differences between conservatives (*M* =5.00, *SE*= .08) and liberals (*M* =4.97, *SE=*.08; *Mdiff*= .03, *t*(483)=.26, *p* = .994).

## 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. We expected that i) the good changes had a positive score; ii) bad change had a negative score; iii) neutral change had a midpoint score. We also conducted a series of one-sample t-tests on moralized vignettes. We summarized descriptives and one-sample t-test results for Study 1 in Table 12.

Table 12

*Study 1: Morality valence check*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Item | Good change | | | Bad change | | |
| *M (SD)* | *t* | *p* | *M (SD)* | *t* | *p* |
| Alcoholism | -1.49 (59.5) | -.56 | .575 | 1.36 (58.0) | .53 | .599 |
| Boss | -1.59 (61.2) | -.58 | .562 | .10 (58.3) | .04 | .968 |
| Parent | -2.47 (59.6) | -.93 | .355 | 1.38 (58.0) | .53 | .594 |
| Terrorism | -1.34 (62.9) | -.48 | .635 | 1.69 (57.6) | .66 | .511 |
| Ethnic Minorities | 4.67 (56.5) | 1.85 | .065 | .76 (57.8) | .29 | .594 |
| Businessman | -2.32 (57.6) | -.90 | .369 | -1.32 (57.6) | -.51 | .609 |
| Romantic  Partner | .22 (57.6) | .08 | .933 | -1.62 (59.0) | -.61 | .540 |
| Police Officer | 3.77 (59.3) | 1.42 | .156 | -.34 (56.5) | -.13 | .893 |
|  | Behavior block 1 (left) | | | Behavior block 2 (right) | | |
| Mac/PC | .42 (56.0) |  |  | -2.86(59.2) |  |  |
| Country/City | .56 (57.8) |  |  | -1.36(60.0) |  |  |
| Cat/Dog | -1.66 (58.7) |  |  | -3.08(57.7) |  |  |
| Football/  Baseball | 2.88 (57.5) |  |  | -.87 (57.0) |  |  |

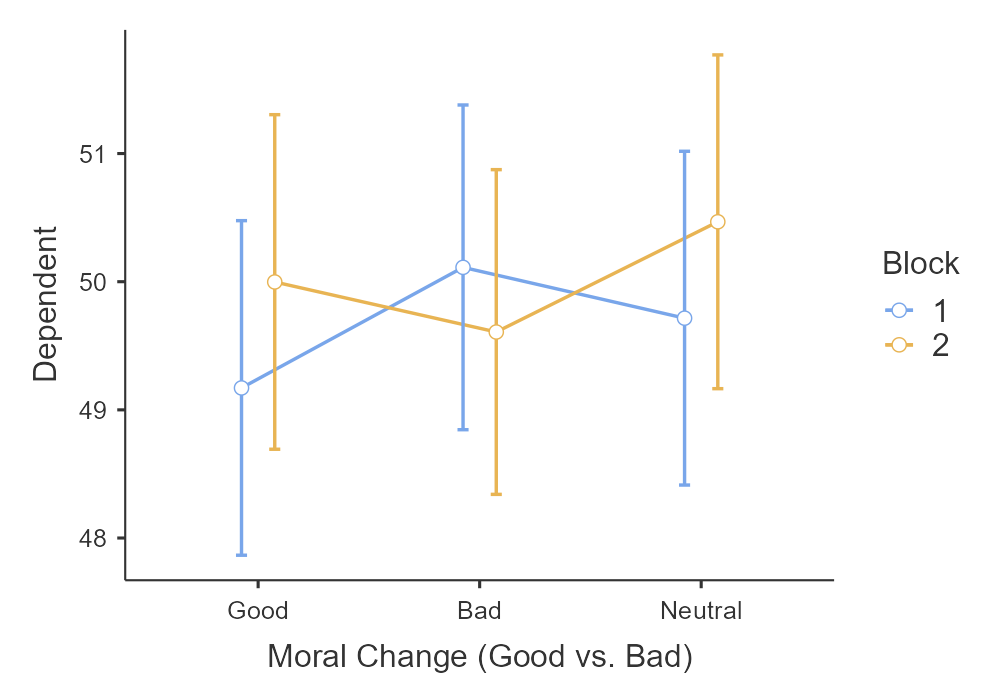
*Note*. M indicates mean. SD indicates standard deviation. *N* = 1000. “Behavior block 1 (left)” and “Behavior block 2 (right)” refers to sets of neutral items, where block refers to the block of display and right/left refer to which of the pair is displayed. For example, in the Mac/PC pair, Mac = Behavior block 1 (left), PC = Behavior block 2 (right).

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

For the true self measure, we conducted a 3 (moral valence: good, bad, neutral; within) x 2 (order: block 1 and block 2; between) mixed-model ANOVA and found no support for a main effect of vignette type (*F*(2,1996) = .31, *p* = .737, *η*²= .00, 95%CI [.00, .01]), block type, or interaction. We found no support for differences between true self ratings for good change (*M* = 49.6, *SD* = 14.9) versus bad change (*M* = 49.9, *SD* = 14.4); *M*diff = -.28, *SE*=.63, *t*(998) = -.43, *p* = .901) (figure 3). Similarly, we found no support when compared to neutral change (*M* = 50.1, *SD* = 14.8; *M*diff = .-.51, *SE* =.67, *t*(998) = -.76, *p* = .727).

Figure 3

*Study 1: Continuous true self measure on moralized changes*

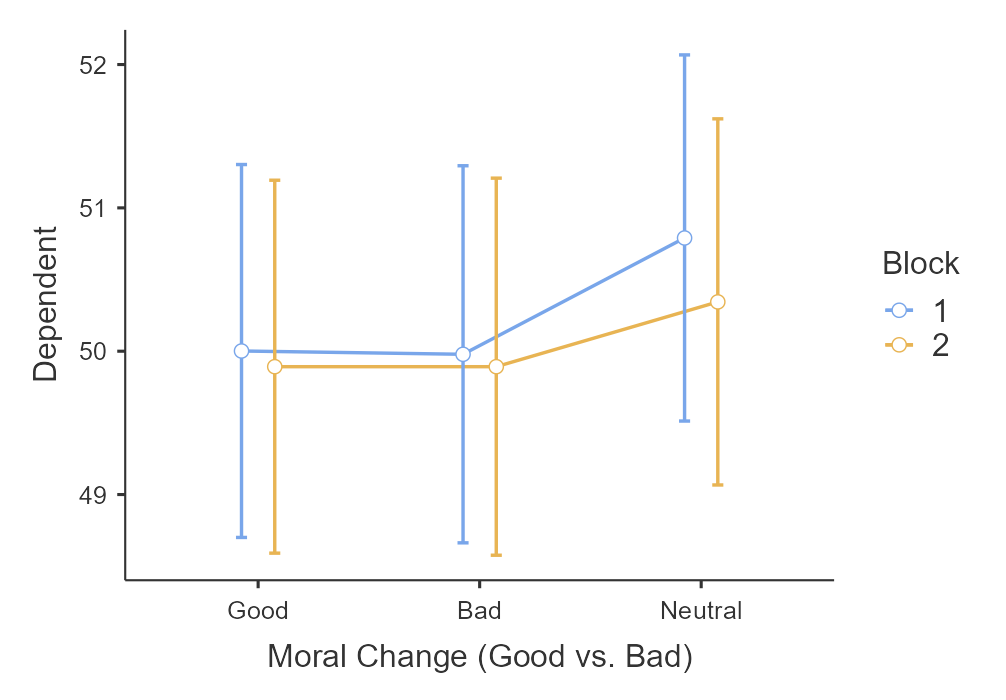


*Note.* Continuous true self measure on good, bad and neutral changes in block 1 and block 2.

For the surface self measure, we conducted another 3 (moral valence: good, bad, neutral; within) x 2 (order: block 1 and block 2; between) mixed-model ANOVA and found no support for a main effect of vignette type (*F*(2,1996) = .71, *p* = .491, *η*²= .00, 95%CI [.00, .01]), block type, or interaction. We found no support for differences between surface self ratings for good change (*M* = 49.9, *SD* = 14.8) versus bad change (*M* = 49.9, *SD* = 15.0); *M*diff = .01, *SE*=.48, *t*(998) = .02, *p* = 1.000) (figure 4). Similarly we found no support when comparing to neutral change (*M* = 50.6, *SD* = 14.5; *M*diff = -.62, *SE* =.65, *t*(998) = -.95, *p* = .609). We failed to support the hypothesis that morally good is less likely to reveal the surface self.

Figure 4

*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 changes in Study 2 aligned with the target article authors’ hypothesized political affiliation with liberal or conservatives. We expected that i) the liberal change had a positive score; ii) conservative change had a negative score. We conducted another series of one-sample t-tests on each vignette. We summarized the descriptives and one-sample t-test results for Study 2 in Table 13.

Table 13

*Study2 : Vignette political view manipulation check*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Item | *M (SD)* | *t* | *p* | |  |
|  | Conservative change |  |  | |  | |
| Homosexuality | | 1.3(56.9) | .72 | .471 | |  |
| Patriotism | | -1.05(58.9) | -.56 | .571 | | |
| Theism | | 5.44(59.1) | 2.91 | .004 | |  |
| Monogamy | | -3.34(58.2) | -1.81 | .070 | | |
|  | Liberal change |  |  | |  | |
| Global warming | | -2.17(58.9) | -1.17 | .243 | |  |
| Gender equality | | 2.83(58.3) | 1.53 | .126 | | |
| Financial success | | 2.61(57.3) | 1.44 | .149 | |  |
| Abortion | | .73(58.2) | .39 | .693 | |  |

*Note*. M indicates mean. SD indicates standard deviation. *N* = 1000

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

We conducted a 2 (item types: liberal and conservative; within) repeated ANOVA with a continuous covariate of political orientation measure. We found no support for an interaction effect (*F*(1,998) = .03, *p* = .856, *η*²= .00, 95% CI [.00, .00]) or between subject effect. The result is (similar to /different from) the interaction effect using the dichotomous political orientation measure.

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

First, we tested the associations between perceived social norms and morality in Studies 1 and 2. There was no support for the associations between social norms and morality in both studies. We provided the results for the specific vignettes in Studies 1 and 2 in the “Additional tables and figures” subsection of supplementary materials.

Second, we tested associations between social norms and true self attributions. We found no support for the link between social norms and overall true self attributions in both Studies 1 and 2 (Table 14). We provided the results for the specific vignettes in Studies 1 and 2in the “Additional tables and figures” subsection of supplementary materials.

Table 14

*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 | Good change | .04 | .163 | -.02, .11 |
|  |  | Bad change | .03 | .356 | -.03, .09 |
| 1 | Continuous true self rating | Good change | .01 | .732 | -.05, .07 |
|  |  | Bad change | .02 | .445 | -.04, .09 |
| 2 | Continuous true self rating | Liberal change | -.00 | .894 | -.06, .06 |
|  |  | Conservative change | .01 | .813 | -.06, .07 |
| Extension | | | | | |
| 1 | Continuous true self and surface self measure | Good change | .04 | .170 | -.02, .11 |
|  |  | Bad change | .01 | .686 | -.05, .08 |

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

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

We conducted a correlational analysis exploring the associations between intuition and true self attributions in Studies 1 and 2. In Table 15 we summarized a comparison of the true self belief of others and true self attributions. In Table 16 we summarized a comparison of the true self belief of one's self and true self attributions.

We found no support for the link between respondents’ own true self belief and true self attributions; we found no support for the link between true self belief of others and true self attributions . Further, we found (no/weak) support for the link between true self rating of their own self and true self rating of others *(r*=-.02, *p*= .555; *95%CI* [-.08, .04]) . No link between surface self rating of their own surface self and others’ surface self *(r*= -.02, *p*=.573; *95% CI* [-.08, .04]).

## Comparing replication to original findings

The interpretation of outcome will be based on LeBel et al. (2019)’s criteria. At this stage we only had a dummy dataset generated. We will compare the replication with the original study when we have actual data.

Table 15

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

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Positive belief of the true self of others** | | | **Negative belief of the true self of others** | | |
| S | Items | *r* | *p* | CI | *r* | *p* | CI |
| 1 | Forced-choice measure (replication) | | | | | | |
| Good change  Bad change | -.02 | .926 | [-.06, .05] | .04 | .180 | [-.02, .10] |
| .01 | .603 | [-.04, .07] | -.00 | .851 | [-.06, .05] |
| 1 | Continuous true self rating (replication) | | | | | | |
| Good change  Bad change | .00 | .892 | [-.05, .06] | -.07 | .028 | [-.13, -.00] |
| -.02 | .479 | [-.08, .04] | -.07 | .022 | [-.13, -.01] |
| 2 | Continuous true self rating (replication) | | | | | | |
| Liberal change  Conservative change | .01 | .735 | [-.05, .07] | .01 | .757 | [-.05, .07] |
| .05 | .072 | [-.00, .11] | .02 | .427 | [-.03,.08] |
| 1 | Continuous true self measure (extension) | | | | | | |
| Good change  Bad change | -.05 | .110 | [-.11, .01] | -.04 | .179 | [-10, .02] |
| .04 | .134 | [-.01, .10] | .02 | .535 | [-.04, .08] |

*.*

Table 16

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

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Positive belief of own true self** | | | **Negative belief of own true self** | | |
| S | Items | *r* | *p* | CI | *r* | *p* | CI |
| 1 | Forced-choice measure (replication) | | | | | | |
| Good change  Bad change | .00 | .900 | [-.05, .06] | .03 | .322 | [-.03, .09] |
| -.03 | .308 | [-.09, .03] | -.04 | .139 | [-.10, .01] |
| 1 | Continuous true self rating (replication) | | | | | | |
| Good change  Bad change | .00 | .951 | [-.06, .06] | .01 | .551 | [-.04, .08] |
| -.01 | .747 | [-.07, .05] | .01 | .551 | [-.04,.08] |
| 2 | Continuous true self rating (replication) | | | | | | |
| Liberal change  Conservative change | .01 | .751 | [-.05, .07] | -.05 | .111 | [-.11, .01] |
| -.00 | .927 | [-.06,.05] | -.00 | .942 | [-.06,.06] |
| 1 | Continuous true self measure (extension) | | | | | | |
| Good change  Bad change | -.01 | .739 | [-.07, .05] | -.01 | .568 | [-.02, .04] |
| .00 | .921 | [-.05,.06] | .04 | .199 | [-.02,.10] |

# Discussion

## Extension: Perceived social norms

[Given feedback from reviewers, we will follow up with a discussion of the findings found in our exploratory direction and the competing hypotheses on the associations between social norms, morality, and true-self perceptions.

## Limitations and future directions

[We will comment on the issues of: Limitations regarding claims of causality, and the need to follow up with experimental work to determine causality.]

[Comments made by the reviewers as potential points for discussion: The forced choice measure of true self in Study 1 can be improved . “ This person’s “surface self” (the things this person learned from society or others)” is not neutral enough to capture surface self as the item seems to suggest that the learnt thought or behaviors are excluded from true self. ]

[In addition, addressing phrasing as “essential” or “non-essential”. For example, rephrasing the forced-choice measure 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) (credits to Dr. Christy A.G). ]

[Future direction to explore the effect of true self across actions with different intentions (credits to Dr. Sergio Barbosa). ]

# References

Adelina, N., & Feldman, G. (2021). Are past and future selves perceived differently from present self? Replication and extension of Pronin and Ross (2006) temporal differences in trait self-ascriptions. *International Review of Social Psychology*, 34(1): 29, 1–16. DOI: 10.5334/irsp.571

Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F., Geller, J., Giner-Sorolla, R., Grange, J. A., Perugini, M., Spies, J., & Van 't Veer, A. (2013). The replication recipe: What makes for a convincing replication? *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2283856>

Berent, I., & Platt, M. (2021). The true “me”—mind or body? *Journal of Experimental Social Psychology*, *93*, 104100. <https://doi.org/10.1016/j.jesp.2020.104100>

Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., ... & De Rosario, M. H. (2018). Package ‘pwr’.

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.

Estimating the reproducibility of psychological science. (2015). *Science*, *349*(6251). https://doi.org/10.1126/science.aac4716

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.

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

Heirene, R. (2020, January 7). A call for replications of addiction research: Which studies should we replicate & what constitutes a “successful” replication?. https://doi.org/10.31234/osf.io/xzmn4

Heiphetz, L. (2019). Moral essentialism and generosity among children and adults. *Journal of   
Experimental Psychology: General*, *148*(12), 2077–2090. https://doi.org/10.1037/xge0000587

Nanakdewa, K., Bulchand, D., Chen, J., Chia, R. J., Lim, V., Ong, C. W., ... & Feldman, G. (2022) Outcomes associated with believing in free will: Meta-analysis Registered Report. *Journal of Research in Personality*. In-principle acceptance of Stage 1 Registered Report. Retrieved from: <https://osf.io/57sau/>

Newman, G. E., Bloom, P., & Knobe, J. (2014). Value judgments and the true self. *Personality and Social Psychology Bulletin*, *40*(2), 203–216. https://doi.org/10.1177/0146167213508791

Newman, G. E., De Freitas, J., & Knobe, J. (2014). 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

Nosek, B. A., & Errington, T. M. (2020). What is replication?. *PLOS Biology*, 18(3), e3000691. https://doi.org/10.1371/journal.pbio.3000691

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

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. (2013). The essential moral self. *PsycEXTRA Dataset*.   
 <https://doi.org/10.1037/e505052014-054>

Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology—a discussion and   
 suggested template. <https://doi.org/10.31234/osf.io/4frms>

Vonasch, A., Hung, W., Leung, W., Nguyen, T., Chan, S., Cheng, B., & Feldman‎, G. (2022). "Less is better" in separate evaluations versus "More is better" in joint evaluations: Mostly successful close replication and extension of Hsee (1998). DOI 10.17605/OSF.IO/9UWNS, retrieved from <https://osf.io/nhyp9/>

Yeung, S. & Feldman, G. (2022). Revisiting the Temporal Pattern of Regret: Replication of Gilovich and Medvec (1994) with extensions examining responsibility. DOI 10.17605/OSF.IO/7M3Q2, retrieved from <https://osf.io/vncy7/>

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, B. (2017). Making replication mainstream.   
 https://doi.org/10.31234/osf.io/4tg9c