Replication and Extension of Alicke (1985) Better-Than-Average Effect for Desirable and Controllable Traits

Ignazio Ziano[1], Pui Yan (Cora) Mok[2], and Gilad Feldman[2] $\textcircled{\scriptsize D}$

Abstract

People tend to regard themselves as better than average. We conducted a replication and extension of Alicke's classic study on trait dimensions in evaluations of self versus others with U.S. American Mechanical Turk workers in two waves (total $N = 1,573$; 149 total traits). We successfully replicated the trait desirability effect, such that participants rated more desirable traits as being more descriptive of themselves than of others (original: $\eta_p^2 = .78$, 95% confidence interval [CI] [.73, .81]; replication: $sr^2 = .54$, 95% CI [.43, .65]). The effect of desirability was stronger for more controllable traits (effect of Desirability × Controllability interaction on self–other-ratings difference; original: $\eta_p^2 = .21$, 95% CI [.12, .28]; replication: $sr^2 = .07$, 95% CI [.02, .12]). In an extension, we found that desirable traits were rated as more common for others, but not for the self. Thirty-five years later, the better-than-average effect appears to remain robust. All materials, data, and code are available at https://osf.io/2y6wj/.

Keywords

better-than-average effect, self-evaluation, comparative judgment, replication

People seem to regard themselves as better than average in many domains. When asked to compare themselves with the average other, people tend to rate themselves possessing more positive traits, being better drivers, and engaging in more desirable behaviors such as contributing to charity (Brown, 2012; Epley & Dunning, 2000; Svenson, 1981). This better-than-average effect—the tendency to evaluate oneself more favorably than the average other person—has received wide attention in the social psychology literature (Alicke & Govorun, 2005; Krueger & Mueller, 2002).

The better-than-average effect has implications for human decision making and judgment. People often make decisions based on how they view themselves in comparison to the average other person. Such self-evaluation may concern their skills, personal attributes, or even physical conditions thus influencing many domains of life including education, health, business, and sports (Dunning et al., 2004; Guenther et al., 2015; Malmendier & Tate, 2005; Stanley et al., 2017; Taylor & Brown, 1988; Ziano & Villanova, 2020). If their evaluation is indeed inaccurate, it is necessary to understand the process behind the phenomenon.

There are two types of explanations for the better-than-average effect. The motivational explanation argues that the phenomenon is a type of self-enhancement for people to protect and maintain their self-worth (Alicke et al., 2013; Sedikides et al., 2003). On the other hand, the nonmotivational explanation suggests the better-than-average effect arises from biases

in information processing. It may be easier for people to evaluate a single object than an abstract entity like the average other, which can lead to inaccurate comparative judgment in the better-than-average paradigm (Chambers & Windschitl, 2004; Krizan & Suls, 2008), and the vagueness of the scale may also play a part (Logg et al., 2018), such that better-than-average effects are stronger when the scale is vague and leaves some space for arbitrary interpretation compared to when the scale is more concrete. While both interpretations may be relevant, researchers have yet to identify a more parsimonious explanation that reconciles them.

Choice of Study for Replication

We aimed to conduct a direct replication of Alicke (1985), one of the classic studies on the better-than-average effect. We selected Alicke for several reasons. One is its academic impact. Published more than 3 decades ago, the study is one of the earliest attempts to demonstrate the better-than-average effect. At the time of writing, it had more than 1,100 citations according

[1] Grenoble Ecole de Management, Grenoble, France
[2] Department of Psychology, The University of Hong Kong, Hong Kong

Corresponding Author:
Gilad Feldman, Department of Psychology, The University of Hong Kong, Pokfulam Road, Hong Kong.
Email: gfeldman@hku.hk; giladfel@gmail.com

Table 1. Summary of Replication and Extension Hypotheses.

Replication Hypotheses
1. The difference between evaluation of self and others is higher as traits increase in desirability.
2. Among high desirable traits, self-ratings are higher than other-ratings for high controllable traits than for low-controllable traits, whereas among low-desirable traits, self-ratings are higher than other-ratings for low-controllable traits than for high-controllable traits.

Extension Hypothesis
3. For ratings of others, trait desirability is positively associated with trait commonness. For ratings of self, trait desirability is negatively associated with trait commonness.

---

to Google Scholar, including those by prominent review papers and textbooks (Brown, 2014; Dunning et al., 2004; Mischel et al., 2007; Taylor & Brown, 1988). Second, to the best of our knowledge, there are no direct replications but only conceptual replications of the study. Building on the findings from Alicke, some studies have found support for the better-than-average effect such that people tended to regard more positive traits as more descriptive of themselves than of the average other (Brown, 2012; Kanten & Teigen, 2008; Pedregon et al., 2012). However, conceptual replications alone cannot verify the robustness of the original findings (Simmons et al., 2011) as differences in procedure and stimuli could cause discrepant results. Direct replications can fill this gap. By operationalizing the variables in the same way as the original, they may help retest these findings and examine whether they are solid foundations for building and strengthening theories (Zwaan et al., 2018).

Alicke (1985) conducted two data collections. In the first, he asked undergraduates to rate various traits in terms of their desirability and controllability, and these were used to form categories of desirability and controllability for an experiment. Participants in a second sample were asked to rate how well these traits characterize them and the average college student.

Alicke's (1985) findings revealed that participants rated more desirable traits as more characteristic of themselves than the average student. Further, when the traits were more desirable, participants believed that traits of higher controllability were even more characteristic of themselves compared to others, generating a desirability by controllability interaction on self-minus-other ratings.

Replication and Extension

We planned to revisit the original findings with two replication hypotheses and extend the original article with one extension hypothesis. We summarized the hypotheses in the present study in Table 1.

The extension investigated the role of desirability and rating perspective on commonness (i.e., how widespread in the population the trait is perceived to be) with the goal to address the methodological concerns around the better-than-average effect. Some researchers have argued that the effect may result from the ambiguous criteria involved in comparative judgments between the self and the average other person (Dunning et al., 1989). Such ambiguity might leave the criteria open to participants' interpretation and thus confound results. A similar argument addressing Alicke's (1985) study was that it may have confounded trait desirability with commonness (Moore, 2007). It suggested that people might report perceiving traits like friendliness as more self-descriptive as these traits were likely displayed more often than those like rudeness or dishonesty. In this regard, trait desirability may be confounded with trait commonness. On the other hand, research has suggested that people with higher self-esteem perceive their desirable traits as less common (Ditto & Griffin, 1993). If people are motivated to enhance their self-view as the better-than-average effect would postulate, there is a reason to believe that the relationship between desirability and commonness may be dependent on the rating perspective, such that they would find desirable traits with higher self-ratings less common and those with higher other-ratings more common.

Adjustments to the Original Study

We made four adjustments to the original procedure. First, we changed the design of the second data collection. After completing the first data collection, we conducted an initial analysis of the results with the goal to categorize traits into four levels of desirability and two levels of controllability. However, results from the first sample prompted a departure from this plan. Only slight decimal differences were observed in the ratings of desirability and controllability, which would pose a challenge to categorizing the variables into meaningful levels. Additionally, categorizing continuous predictors may weaken the ability to detect actual relationships (Irwin & McClelland, 2003; MacCallum et al., 2002). To examine the relationships between the variables more meaningfully, we decided to change the second data collection. Instead of assigning participants to specific levels of desirability and controllability, we randomly assigned participants to one of the three conditions: ratings from the self-perspective, from the average American perspective, or in terms of trait commonness.

Second, whereas in the original article, the same participants rated themselves and the average students (a within-subjects design for the self–other ratings), we had one group of participants rating themselves and another group rating the average American (a between-subjects design).

Third, we conducted the final analysis on an item level. From all participant ratings, we calculated the mean for each trait on each dimension. To validate the change in our planned analysis, we tested the item-level analysis using the data obtained in the first data collection and on a randomly simulated data set on the planned for the second data collection.

Fourth, we addressed the rating perspective by examining the effects of desirability and controllability on self-minus-other ratings, instead of treating it as a predictor. The decision to examine only two-way interactions aimed to improve the clarity of interpretation in statistical analyses. The change

helped address the issue of drawing inferences about the relative importance of multiple, two-way interactions in a complex, three-way analysis of variance, given their differential levels of residual variance (McClelland & Judd, 1993).

Preregistration and Open Science

For the two data collections, we first preregistered the experiment on the Open Science Framework (OSF) and data collection was launched soon after. Preregistrations, disclosures, power analyses, and all materials are available in the Supplementary Material. These, together with data sets and R/RMarkdown code, were made available on the OSF at https://osf.io/2y6wj/.

All measures, manipulations, and exclusions for this investigation are reported, and data collection was completed before analyses. Preregistrations are all available on the OSF:[1] First data collection preregistration (https://osf.io/fyzwd); updated second data collection preregistration following first data collection insights (https://osf.io/9esva).

Method

Power Analysis

We preregistered a power analysis of the results described in Alicke (1985) and included the analysis in the Supplementary Materials ($\alpha = .05$, one-tailed, power $= .95$; G*Power Version 3.1.9.3). Based on the original reported test statistics of the Desirability $\times$ Controllability interaction, $F(3, 261) = 22.72$, we calculated an effect size estimate of $\eta_p^2 = .21$, 95% CI [.12, .28]. With this estimate, a minimum of 71 participants were required to achieve 95% power with an α level of .05 for each condition. Having also taken into account the switch to a between-subjects design, which typically has lower statistical power than within-subjects designs, we aimed for at least 640 participants for the initial ratings and at least 894 participants for the second sample.

Participants and Procedure

Both the first and second samples were recruited online via Amazon Mechanical Turk (MTurk) using Cloud Research (Litman et al., 2017) in return for $0.50 (estimated completion time ~ 4 min, to meet minimum federal wage of U.S. $7.25). The first sample comprised of 670 participants who rated the degree of desirability ($n = 329$) or controllability ($n = 341$) of traits to the average American. The second sample comprised of 903 participants who rated the degree to which these traits characterized themselves ($n = 300$) or the average American ($n = 306$), or the degree of commonness of these traits to the average American ($n = 297$). Six participants, four in the first sample and two in the second sample, were excluded from the analyses since they were detected to be based outside the United States and therefore were not allowed to proceed and answer the questionnaire. A comparison between the study characteristics between the original study and the replication is summarized in Tables S11 and S12 in the Supplementary Materials.

We did not have access to an American undergraduate student population (recruited for the original study) for this replication. We used MTurk samples because of the convenience MTurk provides in reaching a large enough sample size in a short time. MTurk samples have been shown to produce very similar results to U.S. representative samples in experimental political psychology (Coppock, 2017; Coppock et al., 2018; Mullinix et al., 2015). Coming to social psychology results, there are several examples of replication of studies originally conducted with U.S. American undergraduate students who were successfully replicated with MTurk. For instance, overestimation of others' willingness to pay (Frederick, 2012) was successfully replicated on MTurk (Jung et al., 2019; Study 3). An ongoing mass-replication effort successfully replicated a large number of judgment and decision-making studies using Amazon MTurk, with results consistent with student samples and other online recruitment platforms such as Prolific (Chandrashekar et al., 2019; Chen et al., 2020; Collaborative Open-Science Research, 2020; Ziano et al., 2020). Overall, this supports MTurk as a viable sample for replication of Alicke (1985).

The surveys for both the first and second samples were conducted online using Qualtrics. Participants were randomly assigned to one of the conditions. They then received instructions about the rating criteria of their assigned condition and answered comprehension questions accordingly. After answering these questions, they were asked to evaluate 40 traits randomized out of the 149 traits derived from Alicke's (1985) study.

Design and Analyses

The present study is a between-subjects design with three independent variables (IVs; trait desirability, commonness, and controllability) and two dependent variables (DVs; self-ratings and other-ratings). Analyses were conducted on an item level by averaging all participant ratings on each dimension for each trait.

To account for the rating perspective in the replication hypotheses, we calculated self-minus-other ratings by subtracting other-ratings from self-ratings. A positive value indicates that participants perceived the specific trait as more characteristic of themselves than of the average other, whereas a negative value indicates that the trait is regarded as less characteristic of themselves than of the average other. For both data collections, details about attention checks and exclusion criteria are available in the Supplementary Materials.

Materials

First data collection. Before being able to proceed with the survey, participants were asked three comprehension checks (described in detail in the Supplementary Materials).

Desirability. A desirable characteristic was defined as something the average American perceives as good to have and an undesirable characteristic as something the average American perceives as bad to have. This definition was identical to that in Alicke (1985), except that the original reference point "average college student" was replaced by "average American" in order to cater to the participant population in the present study. Participants rated to what extent a trait was desirable (1 = *very undesirable*; 7 = *very desirable*).

Controllability. A controllable characteristic was defined as something that an average American can create or eliminate with a sufficient amount of effort, whereas an uncontrollable characteristic was something that an average American's effort would not suffice to create or eliminate. This definition was identical to that in Alicke (1985), except that the original reference point "average college student" was replaced by "average American" in order to cater to the participant population in the present study. Participants rated to what extent a trait was controllable on a scale from 1 to 7 (1 = *very uncontrollable*; 7 = *very controllable*).

Second Data Collection. Before being able to proceed with the survey, participants were asked three comprehension checks (described in detail in the Supplementary Materials).

Commonness. A common characteristic was defined as one that an average American frequently displays, whereas an uncommon characteristic was defined as something that an average American rarely displays. This definition of commonness was taken from Moore's (2007) review paper, which argues that this dimension was a potential confound with desirability in the original study. Participants rated to what extent a trait was common (1 = *very uncommon*; 7 = *very common*).

Traits. A total of 149 traits were used in the present study. These traits were originally derived from Anderson (1968) and are identical to the final list reported in the appendix of Alicke (1985). Although the study reported using 154 traits, a detailed examination of the list revealed a total number of only 149. In the present study, participants in each condition rated 40 of these traits in randomized order. These traits are summarized in the Supplementary Material.

For self-ratings, participants rated to what extent a trait characterized themselves (1 = *not at all characteristic of me*; 7 = *very characteristic of me*). For others' ratings, participants rated to what extent a trait characterized the average American (1 = *not at all characteristic of the average American*; 7 = *very characteristic of the average American*).

Classification of Replication

The replication was identical to the original in terms of the operationalization and stimuli used for both the IV and the DV. It differed from the original in the procedural details, physical settings, and contextual variables. According to LeBel

Table 2. Classification of Replication Based on LeBel et al.'s (2018) Taxonomy.

Design Facet	Replication
Independent variable operationalization	Same
Dependent variable operationalization	Same
Independent variable stimuli	Same
Dependent variable stimuli	Same
Procedural details	Different
Physical settings	Different
Contextual variables	Different

et al.'s (2018) taxonomy, the present study meets the criteria for a close replication (see Table 2).

Results

We summarized means, standard deviations (*SD*s), and correlations in Table 3 and the means and *SD*s of each dimension for all traits in Table 4. To investigate the relationships between the types of ratings, we performed correlation analyses, correlation comparisons, and multiple linear regression analyses on the item level. In regression analyses, all variables were centered on calculating the interaction term to avoid the problem of multicollinearity (Aiken et al., 1991). The significance level was defined by $p < .05$, one-tailed test for replication hypotheses and two-tailed test for the extension hypothesis. To determine the relative magnitude of each predictor, we used squared semi-partial correlation coefficient to address unique variance explained by the specific predictor when holding other predictors constant. In line with the original study, we calculated an additional DV (self–other ratings) by subtracting other-ratings from self-ratings.

Replication

We found strong support for the desirability effect hypothesis. We ran a correlation analysis and found that desirability had a positive association with self-minus-other ratings ($r = .77$, 95% CI [.69, .82], $p < .001$). This relationship is illustrated in Figure 1. We found support for differences between the desirability self-ratings correlation ($r = .92$, 95% CI [.89, .94], $p < .001$) and the desirability and other-ratings correlation ($r = .61$, 95% CI [.50, .70], $p < .001$; comparison: $z = 8.76$, $p < .001$).

We conclude that regardless of the rating perspective, participants perceived more desirable traits as more descriptive of themselves or the average other, and this positive relationship was stronger for self-ratings than for other-ratings. We proceeded to conduct a multiple linear regression analysis to investigate whether desirability and controllability interacted to predict self-minus-other-ratings and summarized findings in Table 5. First, desirability and controllability were entered into the model. We found that the overall regression was statistically significant, $r = .77$, $R^2 = .59$, 95% CI [.48, .66], $F(2, 146) = 104.30$, $p < .00$. Next, the interaction term was added to the model, which accounted for variance in self-minus-

Table 3. Means, Standard Deviations, and Correlations With Confidence Intervals.

Variable	M	SD	Desirability	Controllability	Commonness	Self-Ratings	Other-Ratings
Desirability	3.73	1.78					
Controllability	4.86	0.81	.03 [−.13, .19]				
Commonness	4.07	0.54	.64** [.54, .73]	.22** [.06, .37]			
Self-ratings	3.73	1.28	.92** [.89, .94]	.04 [−.12, .20]	.61** [.50, .70]		
Other-ratings	3.97	0.58	.61** [.50, .70]	.14 [−.02, .30]	.92** [.89, .94]	.55** [.42, .65]	
Self-minus-other ratings	−0.24	1.08	.77** [.69, .82]	−.03 [−.19, .13]	.23** [.07, .37]	.89** [.86, .92]	.11 [−.05, .27]

Note. M and SD are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. Analyses were conducted on an item level. Ratings of desirability and controllability were collected in the first sample and those of commonness, self-ratings, and other-ratings were collected in the second sample. Self-minus-other represents self-ratings deducted by other-ratings.
*p < .05. **p < .01.

other ratings, $\Delta R^2 = .07$, 95% CI [.02, .12], $\Delta F(1, 145) = 28.17$, $p < .001$. The relationship between desirability and self-minus-other ratings was moderated by controllability ($b = .19$, $SE = .04$, 95% CI [.12, .26], $p < .001$). The interaction was probed by testing the simple main effects of desirability at two levels of controllability: one SD below the mean and one SD above the mean. We found that an increase in controllability enhanced the positive relationship between desirability and self-minus-other ratings (Figure 2). This means that the higher the trait controllability, the more likely the participants were to regard more desirable traits as more descriptive of themselves than of others.

The effect of desirability was strong, explaining more than half of the variation in self-minus-other ratings when holding controllability and the interaction term constant ($sr^2 = .54$, 95% CI [.43, .65]). The effect of controllability was not statistically significant when controlling for other predictors ($sr^2 = .01$, 95% CI [−.01, .03]). The interaction between desirability and controllability had a greater effect than that of controllability, but noticeably smaller than that of desirability ($sr^2 = .07$, 95% CI [.02, .12]). Results supported the hypothesis that desirability and controllability interact to predict the size of the difference between self-ratings and other-ratings.

Extension: Trait Desirability, Trait Commonness, and Self–Other Ratings

We conducted two multiple linear regression analyses to examine the interaction between self-ratings and desirability, as well as the interaction between other-ratings and desirability on predicting commonness. We summarized the results in Tables S7 and S8 in the Supplementary, respectively. We failed to find support for the relationship between commonness and self or other ratings being dependent on desirability. Our discussion below therefore focuses on the first step of each model when only the two predictors were entered.

Our findings did not fully support the extension hypothesis, as we found support for a relationship between trait desirability and trait commonness, yet found no support for a relationship between trait desirability and self-ratings. Examining desirability and self-ratings, we found support for desirability as predictive of commonness ($b = .16$, $SE = .05$, 95% CI [.07, .26], $p < .001$), but not for self-ratings ($b = .05$, $SE = .07$, 95% CI [−.09, .18], $p = .48$) (see Table S7).

On the other hand, on examining desirability and other-ratings, they were both positively associated with commonness. Other- ratings ($b = .80$, $SE = .04$, 95% CI [.71, .86], $p < .001$) was a stronger predictor than desirability ($b = .04$, $SE = .01$, 95% CI [.01, .06], $p = .001$; see Table S8).

Overall, both regression equations accounted for a significant portion of variance in commonness ratings: $R^2 = .86$, 95% CI [.82, .89], $F(2, 146) = 51.88$, $p < .001$ for desirability and other-ratings and $R^2 = .42$, 95% CI [.29, .51], $F(2, 146) = 451$, $p < .001$ for desirability and self-ratings, respectively.

The association between other-ratings and commonness was the strongest ($sr^2 = .41$, 95% CI [.31, .52]) compared with desirability and the interaction term which accounted for very little variance in commonness in the same model. Desirability was a stronger predictor of commonness ($sr^2 = .04$, 95% CI [−.01, .09]) than self-ratings and the interaction between self-ratings and commonness, which showed very weak effects when entered in the same model.

In our preregistration, we planned to evaluate replication outcomes based on the direction and strength of the detected signals in relation to the original effect size at a 95% CI (LeBel et al., 2018; see Table 6 for a comparison). Given the difference in our statistical analyses from the original, our findings addressed only some of the effects in a different approach and may not be applicable for a direct comparison using LeBel et al.'s (2018) framework. We recommend caution in comparing the effect sizes of Alicke (1985) and of the present replication.

Table 4. Mean Ratings and Standard Deviations of Traits in Terms of Desirability, Controllability, and Commonness and Self-Ratings and Other-Ratings.

Traits	Desirability (N = 149) M (SD)	Controllability (N = 149) M (SD)	Commonness (N = 149) M (SD)	Self (N = 149) M (SD)	Other (N = 149) M (SD)
Intelligent	6.41 (1.02)	4.01 (1.81)	4.57 (1.24)	5.39 (1.35)	4.55 (1.16)
Reliable	6.40 (1.01)	5.93 (1.13)	4.76 (1.31)	5.90 (1.08)	4.57 (1.09)
Loyal	6.33 (1.00)	5.69 (1.42)	4.84 (1.20)	5.93 (1.34)	4.73 (1.10)
Attractive	6.33 (0.97)	3.35 (1.62)	4.32 (1.19)	4.31 (1.39)	4.35 (1.08)
Responsible	6.31 (1.16)	5.82 (1.32)	4.65 (1.20)	5.74 (1.26)	4.67 (1.31)
Resourceful	6.27 (1.01)	5.13 (1.49)	4.8 (1.14)	5.48 (1.12)	4.81 (1.43)
Kind	6.26 (1.08)	5.72 (1.40)	4.8 (1.20)	5.56 (1.19)	4.53 (1.14)
Sincere	6.2 (1.01)	5.33 (1.62)	4.59 (1.14)	5.76 (1.09)	4.51 (1.17)
Friendly	6.16 (1.01)	5.88 (1.19)	4.75 (1.16)	5.51 (1.18)	4.89 (1.35)
Dependable	6.15 (1.33)	5.64 (1.32)	4.78 (1.28)	5.85 (1.27)	4.62 (1.11)
Respectful	6.15 (1.03)	5.97 (1.10)	4.37 (1.37)	5.85 (0.98)	4.15 (1.41)
Admirable	6.12 (1.03)	4.8 (1.68)	4.29 (1.26)	4.38 (1.45)	4.44 (1.36)
Wise	6.09 (1.25)	3.73 (1.62)	4.08 (1.25)	4.85 (1.44)	3.89 (1.34)
Good-tempered	6.09 (1.13)	4.89 (1.70)	4.67 (1.27)	5.10 (1.37)	4.52 (1.28)
Interesting	6.09 (0.97)	4.08 (1.67)	4.37 (1.21)	4.78 (1.34)	4.83 (1.3)
Bright	6.07 (1.00)	4.06 (1.81)	4.45 (1.23)	5.60 (1.21)	4.45 (1.23)
Honorable	6.04 (1.13)	5.30 (1.74)	4.38 (1.28)	5.32 (1.39)	4.49 (1.32)
Clearheaded	6.02 (1.23)	4.58 (1.52)	4.23 (1.28)	5.03 (1.54)	4.11 (1.17)
Pleasant	6.01 (1.20)	5.46 (1.41)	5.03 (1.10)	5.49 (1.07)	4.52 (1.10)
Ethical	6.01 (1.16)	5.58 (1.49)	4.51 (1.14)	5.54 (1.40)	4.33 (1.27)
Levelheaded	5.99 (1.26)	4.71 (1.59)	4.43 (1.24)	5.14 (1.51)	4.32 (1.29)
Intellectual	5.99 (1.16)	4.06 (1.79)	4.19 (1.33)	5.06 (1.17)	3.92 (1.17)
Considerate	5.98 (1.37)	5.99 (1.06)	4.28 (1.41)	5.75 (1.08)	4.47 (1.30)
Self-disciplined	5.98 (1.26)	5.59 (1.24)	4.14 (1.29)	4.70 (1.61)	3.95 (1.32)
Polite	5.98 (1.16)	6.36 (0.87)	4.56 (1.19)	5.65 (1.15)	4.12 (1.36)
Punctual	5.97 (1.04)	6.27 (1.03)	4.41 (1.31)	5.57 (1.56)	4.24 (1.31)
Versatile	5.95 (1.13)	4.87 (1.38)	4.71 (1.17)	4.65 (1.55)	4.55 (1.29)
Clean	5.91 (1.11)	6.20 (1.29)	4.85 (0.95)	5.47 (1.47)	4.68 (1.31)
Humorous	5.82 (1.17)	4.19 (1.54)	4.73 (1.18)	5.22 (1.41)	4.62 (1.10)
Original	5.81 (1.20)	4.02 (1.64)	4.12 (1.26)	4.80 (1.42)	3.98 (1.51)
Grateful	5.80 (1.29)	5.86 (1.34)	4.33 (1.34)	5.30 (1.50)	4.17 (1.51)
Trustful	5.77 (1.24)	5.14 (1.57)	4.78 (1.11)	5.21 (1.46)	4.24 (1.2)
Persistent	5.77 (1.17)	5.68 (1.36)	4.77 (1.23)	5.11 (1.52)	4.79 (1.28)
Lucky	5.76 (1.24)	1.85 (1.45)	4.12 (1.17)	3.43 (1.64)	4.19 (1.30)
Mature	5.71 (1.40)	4.93 (1.70)	4.26 (1.06)	5.32 (1.29)	4.00 (1.14)
Perceptive	5.71 (1.22)	4.03 (1.70)	4.31 (1.16)	5.62 (1.36)	4.27 (1.43)
Sharp-witted	5.71 (1.19)	3.77 (1.81)	4.04 (1.40)	4.92 (1.57)	4.40 (1.26)
Creative	5.64 (1.27)	3.81 (1.63)	4.45 (1.21)	4.79 (1.62)	4.31 (1.45)
Cooperative	5.63 (1.38)	6.17 (1.09)	4.83 (1.34)	5.51 (1.04)	4.57 (1.38)
Observant	5.63 (1.23)	5.42 (1.39)	4.52 (1.22)	5.61 (1.43)	4.29 (1.56)
Lively	5.61 (1.17)	4.57 (1.54)	4.97 (1.21)	4.50 (1.56)	4.91 (0.93)
Clever	5.60 (1.41)	3.89 (1.78)	4.13 (1.38)	5.29 (1.29)	4.22 (1.36)
Imaginative	5.60 (1.26)	3.71 (1.70)	4.61 (1.27)	5.17 (1.41)	4.29 (1.28)
Sportsmanlike	5.55 (1.26)	5.59 (1.47)	4.42 (1.13)	4.67 (1.86)	4.48 (1.25)
Neat	5.54 (1.09)	5.98 (1.16)	4.33 (1.14)	4.84 (1.57)	3.93 (1.28)
Normal	5.46 (1.29)	4.24 (1.58)	5.28 (1.23)	5.12 (1.54)	5.00 (1.12)
Witty	5.40 (1.35)	3.72 (1.71)	4.29 (1.20)	4.69 (1.73)	4.20 (1.31)
Well-read	5.33 (1.34)	5.80 (1.35)	3.84 (1.45)	5.09 (1.51)	3.76 (1.66)
Fearless	5.26 (1.42)	3.81 (1.73)	3.86 (1.53)	3.35 (1.6)	3.99 (1.58)
Bold	5.25 (1.17)	4.79 (1.49)	4.59 (1.13)	3.72 (1.52)	4.73 (1.17)
Quick	5.08 (1.22)	3.94 (1.70)	4.26 (1.19)	4.49 (1.43)	4.31 (1.32)
Fashionable	5.04 (1.16)	5.72 (1.37)	4.46 (1.08)	3.58 (1.84)	4.01 (1.32)
Progressive	5.00 (1.28)	5.28 (1.41)	4.51 (1.25)	4.83 (1.72)	4.40 (1.22)
Ingenious	4.96 (1.79)	3.60 (1.77)	3.69 (1.51)	3.94 (1.75)	3.90 (1.49)
Self-satisfied	4.73 (1.57)	4.81 (1.57)	4.62 (1.38)	3.91 (1.66)	4.81 (1.27)

(continued)

Table 4. (continued)

Traits	Desirability (N = 149) M (SD)	Controllability (N = 149) M (SD)	Commonness (N = 149) M (SD)	Self (N = 149) M (SD)	Other (N = 149) M (SD)
Thrifty	4.70 (1.37)	5.59 (1.48)	3.91 (1.48)	4.96 (1.59)	3.54 (1.36)
Philosophical	4.70 (1.34)	4.48 (1.66)	3.33 (1.45)	4.36 (1.70)	3.43 (1.50)
Prudent	4.67 (1.48)	4.91 (1.50)	4.04 (1.22)	4.38 (1.72)	4.07 (1.42)
Religious	4.62 (1.35)	5.66 (1.56)	4.73 (1.41)	3.21 (2.18)	4.43 (1.31)
Meticulous	4.59 (1.37)	5.26 (1.56)	3.77 (1.06)	4.29 (1.81)	3.70 (1.24)
Obedient	4.52 (1.44)	6.00 (1.16)	4.26 (1.22)	4.45 (1.41)	4.01 (1.44)
Authoritative	4.5 (1.44)	4.91 (1.44)	4.54 (1.21)	3.54 (1.88)	4.22 (1.37)
Changeable	4.36 (1.25)	5.30 (1.62)	4.20 (1.26)	4.04 (1.38)	4.25 (1.38)
Sensitive	4.35 (1.40)	3.92 (1.79)	4.35 (1.22)	4.96 (1.54)	4.42 (1.23)
Conforming	4.11 (1.52)	5.20 (1.41)	4.86 (1.26)	3.94 (1.52)	4.48 (1.37)
Reserved	4.10 (1.38)	4.69 (1.76)	3.72 (1.19)	4.85 (1.70)	3.27 (1.31)
Prideful	4.02 (1.88)	5.35 (1.34)	5.28 (1.21)	3.83 (1.82)	5.26 (1.21)
Impressionable	3.88 (1.69)	4.09 (1.64)	4.65 (1.16)	3.25 (1.53)	4.62 (1.26)
Extravagant	3.77 (1.64)	5.32 (1.72)	4.33 (1.44)	2.76 (1.67)	4.17 (1.46)
Soft-spoken	3.73 (1.40)	4.76 (1.60)	3.19 (1.31)	4.29 (1.75)	3.00 (1.15)
Cunning	3.69 (1.96)	3.90 (1.80)	3.56 (1.31)	3.13 (1.78)	3.91 (1.21)
Choosy	3.53 (1.41)	5.21 (1.32)	4.75 (1.12)	4.35 (1.75)	4.94 (1.23)
Ordinary	3.53 (1.39)	4.06 (1.69)	5.04 (1.22)	4.23 (1.79)	4.60 (1.54)
Eccentric	3.53 (1.30)	4.16 (1.66)	3.44 (1.35)	3.68 (1.91)	3.36 (1.44)
Strict	3.46 (1.43)	5.40 (1.53)	3.42 (1.22)	3.59 (1.85)	3.50 (1.25)
Self-concerned	3.45 (1.58)	5.11 (1.57)	5.08 (1.37)	4.04 (1.50)	5.14 (1.31)
Daydreamer	3.43 (1.31)	4.30 (1.81)	4.41 (1.38)	4.40 (1.78)	4.12 (1.17)
Solemn	3.37 (1.36)	4.92 (1.53)	3.53 (1.35)	3.63 (1.71)	3.23 (1.05)
Overcautious	3.01 (1.22)	4.69 (1.57)	3.83 (1.22)	4.56 (1.84)	3.37 (1.26)
Inhibited	2.94 (1.29)	4.07 (1.65)	3.52 (1.14)	3.30 (1.55)	3.16 (1.44)
Bashful	2.82 (1.44)	3.76 (1.67)	3.24 (1.33)	3.60 (1.95)	3.00 (1.30)
Melancholy	2.76 (1.62)	4.10 (1.63)	3.51 (1.13)	3.19 (1.88)	3.49 (1.30)
Irreligious	2.76 (1.29)	5.17 (1.73)	3.51 (1.43)	3.73 (2.32)	3.51 (1.41)
Impulsive	2.75 (1.30)	4.23 (1.71)	4.27 (1.40)	2.81 (1.50)	5.18 (1.26)
Passive	2.73 (1.29)	4.74 (1.43)	3.76 (1.31)	3.55 (1.71)	3.35 (1.39)
Hesitant	2.70 (1.35)	4.52 (1.62)	3.57 (1.14)	4.03 (1.59)	3.52 (1.27)
Meek	2.62 (1.50)	4.18 (1.72)	2.97 (1.30)	3.35 (1.81)	2.88 (1.10)
Compulsive	2.60 (1.40)	3.93 (1.76)	4.05 (1.54)	2.92 (1.69)	4.42 (1.43)
Restless	2.59 (1.35)	4.30 (1.53)	4.01 (1.52)	3.65 (1.54)	4.36 (1.49)
Boastful	2.58 (1.58)	5.52 (1.55)	4.67 (1.33)	2.33 (1.52)	4.88 (1.08)
Radical	2.51 (1.26)	4.92 (1.67)	3.41 (1.52)	2.71 (1.77)	3.18 (1.54)
Timid	2.47 (1.20)	3.72 (1.89)	3.14 (1.36)	3.52 (1.96)	2.99 (1.19)
Profane	2.42 (1.47)	5.57 (1.51)	4.19 (1.45)	2.76 (1.69)	3.84 (1.47)
Unemotional	2.42 (1.34)	3.61 (1.74)	3.04 (1.42)	2.83 (1.84)	2.55 (1.32)
Unpoised	2.26 (1.28)	4.66 (1.52)	3.68 (1.45)	3.00 (1.70)	3.65 (1.40)
Unoriginal	2.23 (1.31)	3.54 (1.55)	3.94 (1.50)	2.80 (1.70)	3.6 (1.61)
Unsophisticated	2.22 (1.27)	4.49 (1.67)	3.94 (1.43)	2.91 (1.62)	4.08 (1.39)
Discontented	2.15 (1.24)	4.93 (1.66)	4.19 (1.51)	2.76 (1.70)	3.95 (1.51)
Self-centered	2.12 (1.21)	5.28 (1.58)	5.00 (1.25)	2.78 (1.53)	4.98 (1.17)
Humorless	2.09 (1.44)	3.62 (1.87)	2.92 (1.27)	1.94 (1.38)	2.90 (1.40)
Uncultured	2.09 (1.25)	4.71 (1.68)	3.70 (1.39)	2.24 (1.32)	3.74 (1.51)
Unstudious	2.08 (1.36)	5.28 (1.70)	3.62 (1.23)	2.32 (1.48)	3.57 (1.30)
Vain	2.08 (1.17)	4.94 (1.63)	4.43 (1.44)	2.48 (1.67)	4.31 (1.50)
Unforgiving	2.07 (1.22)	5.30 (1.51)	3.62 (1.40)	2.68 (1.60)	3.58 (1.55)
Clumsy	2.02 (1.43)	3.52 (1.80)	3.38 (1.34)	3.40 (1.80)	3.17 (1.36)
Forgetful	2.00 (1.17)	3.59 (1.66)	3.81 (1.27)	3.06 (1.63)	3.76 (1.53)
Unentertaining	1.99 (1.39)	4.04 (1.48)	3.48 (1.36)	2.99 (1.70)	3.17 (1.32)
Cold	1.97 (1.20)	5.01 (1.57)	3.43 (1.33)	2.46 (1.50)	3.16 (1.43)
Withdrawn	1.97 (1.17)	4.43 (1.68)	3.01 (1.33)	3.54 (1.93)	2.93 (1.41)
Gullible	1.93 (1.22)	3.76 (1.67)	4.09 (1.33)	2.85 (1.66)	3.90 (1.48)
Complaining	1.92 (1.41)	5.57 (1.58)	4.65 (1.37)	3.00 (1.61)	4.54 (1.34)

(continued)

Table 4. (continued)

Traits	Desirability (N = 149)	Controllability (N = 149)	Commonness (N = 149)	Self (N = 149)	Other (N = 149)
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
Deceptive	1.91 (1.34)	5.35 (1.68)	3.66 (1.32)	2.17 (1.26)	3.56 (1.42)
Meddlesome	1.91 (1.31)	5.70 (1.35)	3.74 (1.45)	2.21 (1.33)	3.93 (1.28)
Disobedient	1.91 (1.17)	5.82 (1.39)	3.69 (1.49)	2.57 (1.57)	3.69 (1.45)
Maladjusted	1.90 (1.33)	3.86 (1.74)	3.38 (1.48)	2.22 (1.64)	3.17 (1.37)
Dissatisfied	1.90 (1.25)	4.64 (1.68)	4.38 (1.59)	3.14 (1.76)	4.18 (1.51)
Unkind	1.89 (1.32)	5.52 (1.60)	3.30 (1.28)	1.98 (1.32)	3.35 (1.51)
Insecure	1.89 (1.17)	3.75 (1.78)	3.88 (1.55)	3.65 (1.91)	4.01 (1.60)
Irrational	1.87 (1.41)	4.34 (1.74)	3.92 (1.24)	2.26 (1.44)	3.83 (1.31)
Irresponsible	1.85 (1.37)	5.48 (1.57)	3.73 (1.44)	2.32 (1.45)	3.85 (1.42)
Shallow	1.83 (1.14)	4.92 (1.64)	3.99 (1.61)	2.37 (1.59)	4.25 (1.43)
Phony	1.82 (1.41)	5.15 (1.86)	3.58 (1.48)	1.88 (1.37)	3.94 (1.55)
Rude	1.82 (1.36)	5.97 (1.38)	4.00 (1.45)	2.07 (1.45)	3.73 (1.33)
Snobbish	1.81 (1.19)	5.56 (1.61)	3.76 (1.50)	2.14 (1.47)	3.73 (1.53)
Disrespectful	1.80 (1.33)	5.83 (1.46)	3.73 (1.40)	2.02 (1.50)	3.79 (1.59)
Spiteful	1.79 (1.28)	5.14 (1.64)	3.68 (1.46)	2.38 (1.69)	3.6 (1.51)
Uncivil	1.78 (1.28)	5.59 (1.51)	3.22 (1.52)	2.12 (1.61)	3.25 (1.40)
Belligerent	1.78 (1.21)	5.38 (1.52)	3.76 (1.54)	1.98 (1.59)	3.44 (1.52)
Unpopular	1.77 (1.18)	3.46 (1.68)	3.60 (1.35)	3.07 (1.77)	3.44 (1.41)
Unskilled	1.76 (1.16)	4.98 (1.78)	3.64 (1.39)	2.39 (1.51)	3.07 (1.31)
Mean	1.74 (1.17)	5.74 (1.35)	3.53 (1.54)	2.26 (1.62)	3.48 (1.38)
Impolite	1.73 (1.00)	5.98 (1.38)	3.89 (1.42)	2.00 (1.30)	3.73 (1.50)
Unreasonable	1.71 (1.13)	4.76 (1.69)	3.82 (1.36)	1.95 (1.17)	3.63 (1.58)
Tiresome	1.71 (0.99)	4.28 (1.62)	3.67 (1.57)	2.61 (1.70)	3.57 (1.56)
Discourteous	1.70 (1.02)	5.78 (1.56)	3.78 (1.40)	1.90 (1.35)	3.84 (1.49)
Unappreciative	1.66 (1.00)	5.42 (1.75)	4.20 (1.70)	2.04 (1.41)	3.86 (1.48)
Troubled	1.66 (0.95)	3.77 (1.61)	3.92 (1.29)	2.67 (1.65)	3.87 (1.43)
Lazy	1.65 (1.24)	5.59 (1.53)	3.77 (1.43)	2.81 (1.70)	3.78 (1.47)
Ill-mannered	1.65 (1.18)	5.47 (1.69)	3.62 (1.55)	2.15 (1.63)	3.60 (1.60)
Jealous	1.65 (0.94)	4.57 (1.69)	4.28 (1.41)	2.62 (1.56)	3.79 (1.40)
Unpleasant	1.62 (1.18)	5.24 (1.55)	3.44 (1.30)	2.02 (1.36)	3.34 (1.48)
Hostile	1.59 (1.12)	5.35 (1.40)	3.48 (1.57)	1.84 (1.25)	3.28 (1.53)
Deceitful	1.58 (1.02)	5.51 (1.67)	3.70 (1.38)	1.86 (1.37)	3.16 (1.37)
Unethical	1.57 (1.09)	5.43 (1.54)	3.52 (1.57)	1.65 (1.08)	3.41 (1.49)
Liar	1.57 (1.07)	5.98 (1.44)	3.89 (1.68)	2.05 (1.43)	3.33 (1.36)
Dishonorable	1.55 (1.16)	5.22 (1.89)	3.04 (1.50)	1.90 (1.64)	2.88 (1.51)
Unpleasing	1.54 (0.84)	4.49 (1.65)	3.27 (1.22)	2.37 (1.60)	2.89 (1.31)
Incompetent	1.53 (1.26)	4.40 (1.77)	3.30 (1.33)	1.85 (1.37)	3.17 (1.39)
Dishonest	1.53 (1.08)	5.43 (1.67)	3.42 (1.38)	2.05 (1.37)	3.59 (1.27)

Note. The traits are arranged in descending order of desirability ratings.

Discussion

The present study aimed to replicate and extend the findings of Alicke's (1985) study. Alicke found support for the effects of trait dimensions on the difference between self-ratings and other-ratings. In two preregistered data collections, we successfully replicated the effects of desirability, as well as the interaction between desirability and controllability. First, there was a strong positive relationship between trait desirability and the difference between self-ratings and other-ratings on the same trait. The more participants rated a trait as desirable, the more participants rated the trait as more characteristic of themselves than the characteristic of the average American. Second, the effect of desirability on the difference between self-ratings and other-ratings was stronger for highly controllable traits and

weaker (but still positive) for less controllable traits. However, the main effect of controllability was found to be weaker than expected and in the opposite direction to the original (which was significant and positive). Additionally, in our extension, we found that more desirable traits were regarded as more common, yet this only applied to other-ratings, but not self-ratings.

Replication: Effect of Desirability, Controllability, and Their Interaction on Better-Than-Average Effect

In summary, the predictors showed similar *relative* magnitudes as the original study: Desirability showed the strongest effect, followed by the interaction between desirability and controllability, and then controllability. Of particular interest is the
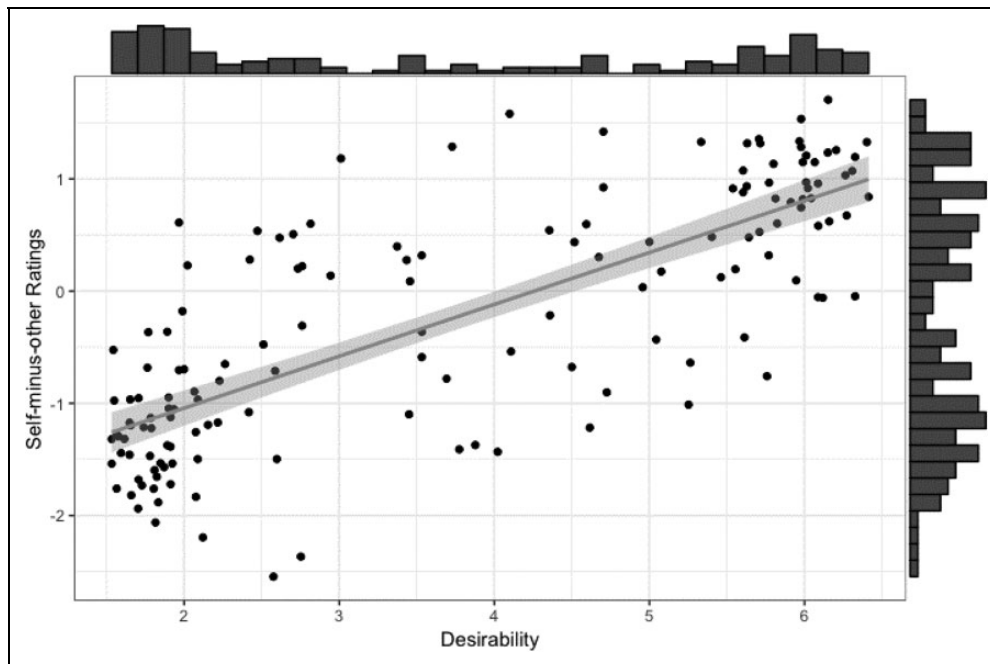
Figure 1. Scatterplot with marginal histograms and 95% confidence interval showing the relationship between desirability and self-minus-other ratings.

Table 5. Regression Results Using Self-Minus-Other Ratings as the Dependent Variable and Desirability and Controllability as the Independent Variables.

Predictor	b	b 95% CI [LL, UL]	β	β 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	Fit	Difference
(Intercept)	−.24***	[−.36, −.13]						
Desirability	.46***	[.40, .53]	0.77	[0.66, 0.87]	.59	[.49, .69]	$R^2 = .59$***	
Controllability	−.07	[−.21, .07]	−0.05	[−0.16, 0.05]	.00	[−.01, .01]	95% CI [.48, .66]	
(Intercept)	−.25***	[−.36, −.15]					$R^2 = .66$***	$\Delta R^2 = .07$***
Desirability	.45***	[.39, .51]	0.74	[0.64, 0.84]	.54	[.43, .65]	95% CI [.56, .71]	95% CI [.02, .12]
							$F(3, 145) = 91.84$***	$\Delta F(1, 145) = 28.17$***
Controllability	−.15*	[−.28, −.02]	−0.11	[−0.21, −0.01]	.01	[−.01, .03]		
Interaction	.19***	[.12, .26]	0.27	[0.17, 0.37]	.07	[.02, .12]		

Note. A significant b weight indicates the β weight and semi-partial correlation are also significant. b represents unstandardized regression weights. β indicates the standardized regression weights. sr^2 represents the semi-partial correlation squared. r represents the zero-order correlation. LL and UL indicate the lower and upper limits of a confidence interval, respectively.
*p < .05. **p < .01. ***p < .001.

consistently strong effect observed in desirability across all analyses in the present study, which appears to support the robustness of the better-than-average effect. Taking into account different trait dimensions, previous studies found that people tended to rate traits as more characteristic of themselves than of the average other when the traits were more important or more positive (Brown, 2012; Pahl & Eiser, 2005; Pedregon et al., 2012). A meta-analysis study suggested that Westerners self-enhanced more than East Asians and that the better-than-average paradigm yielded one of the strongest effects for self-enhancement in both cultures among 31 methods (Heine & Hamamura, 2007). This finding, however, has been

contradicted by subsequent research arguing that there is little difference between Westerners and East Asians in the extent of self-enhancement (Brown, 2010; Zell et al., 2019; also see Ziano et al., 2020; Chandrashekar et al., 2020 for high consistency in findings across American and Hong Kong samples in judgment and decision-making effects).

We replicated the interaction between desirability and controllability. This is in support of related research showing similar moderating effects for controllability, such that controllable traits were regarded as more self-descriptive when described positively, but less so when described negatively (Rothermund et al., 2005). On the other hand, we did not replicate the main
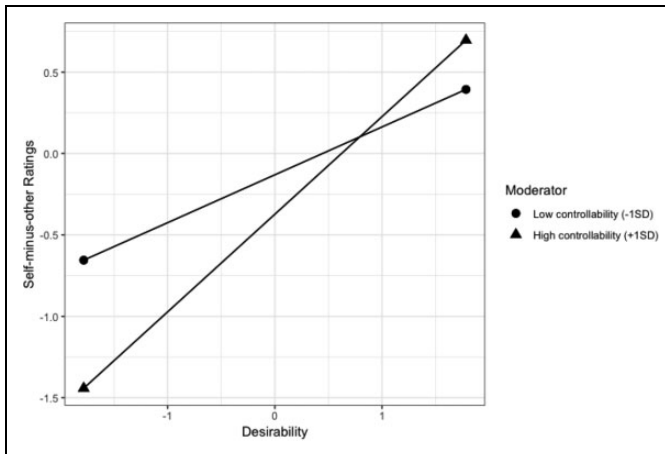
effect of controllability, finding a weak and inconsistent effect, with CIs including the null and the effect in the direction opposite to the original. Given the deviations in both the magnitude and the direction, we consider this finding inconclusive.

Extension: Does Commonness Confound Better-Than-Average Effects?

We found that desirability was positively associated with commonness, yet we found no evidence that this relationship

Figure 2. Simple slopes of desirability predicting self-minus-other ratings for one standard deviation below the mean of controllability and one standard deviation above the mean of controllability.

moderated the self–other ratings difference. The positive association between desirability and commonness was supported by another study by Pahl and Eiser (2005), although authors did not specify how commonness was operationalized. A possible interpretation for this finding is that desirability might be confounded with commonness, as argued by Moore (2007). Specifically, social norms not only determine which traits are desirable but also encourage people to display these traits more often than to display the undesirable ones. As a result, these traits would be regarded as more common.

In comparison with self-ratings, other-ratings were a stronger predictor of commonness. Revisiting the definition of commonness may help explain this finding. In the present study, commonness was operationalized as the degree to which people perceive that a trait is frequently displayed among the average American. Implicated in this definition is observability that may help explain why self-ratings and other-ratings correlated with commonness to varying degrees. For self-evaluation, people can access their inner thoughts and feelings and recall different instances when they display a certain trait. However, evaluating the average American is likely different. Compared with self-evaluation, not only is it more limited to observable traits but it also requires additional cognitive effort of imagining an abstracted average (Chambers & Windschitl, 2004; Krizan & Suls, 2008). In other words, whereas people are aware of both their public and private traits when making self-evaluation, they are likely to base their judgment of the average other on observable traits.

Table 6. Comparison of Effect Sizes Between the Original Article and Replication With the Self-Minus-Other Ratings Difference as Dependent Variable.

Predictor	Original Article		Replication			NHST Summary	Replication Summary
	η_p^2	f	B	β	sr^2		
Desirability	.78 [.73, .81]	1.88 [1.66, 2.06]	.45 [.39, .51]	.74 [.64, .38]	.54 [.43, .65]	Supported	Consistent in direction; strong effect
Controllability	.06 [.002, .18]	0.26 [0.04, 0.47]	−.15 [−.28, −.02]	−.11 [−.21, −.03]	.01 [−.01, .03]	Supported	Inconsistent in direction; inconclusive finding
Desirability × Controllability	.21 [.12, .28]	0.52 [0.37, 0.62]	.19 [.12, .26]	.27 [.17, .37]	.07 [.02, .12]	Supported	Consistent in direction; weak effect
Desirability × Controllability[a]	.15 [.04, .28]	0.42 [0.2, 0.62]					
Desirability × Perspective	.59 [.52, .65]	1.21 [1.04, 1.35]					
Perspective × Desirability × Controllability	.23 [.14, .31]	0.55 [0.40, 0.66]					

Note. Values in square brackets indicate the 95% confidence interval. *b* represents unstandardized regression weights. β indicates the standardized regression weights. sr^2 represents the semi-partial correlation squared. Since this analysis is performed with the self-minus-other ratings as dependent variable, we did not include effects of the Desirability × Perspective and the Perspective × Desirability × Controllability interactions.
[a]The original article revised categorization of desirability at high, neutral-high, neutral-low, and low levels.

Constraints on Generalizability

Sample. We recruited U.S. residents from MTurk as participants. This limits the generalizability of the present results to other populations, especially non-Western, educated, industrialized, rich, democratic ones (Henrich et al., 2010). This is of particular importance given the ongoing controversy in the literature on whether East Asians and Westerners differ in the extent of self-enhancement (Brown, 2010; Heine & Hamamura, 2007; Zell et al., 2019). Research using a different sample may obtain different results.

Materials. We had no access to the original list of traits used for the first sample. The original article mentioned where and how the traits were initially derived, yet the full list was unreported, leading to our decision to use only the 149 traits provided in the original's appendix. It is possible that the perception toward the traits has changed over the past 3 decades and thus different traits would have been shortlisted for the second sample based on the ratings of desirability and controllability. This gap in information calls for more shared documentation in psychological research for facilitating reproducible work. Research using a different list of traits may obtain different results.

"Average American" designation. Note that the designation of "average American" is potentially confusing as "American" can indicate people who are not U.S. citizens (e.g., Bolivians, Mexicans, and Canadians). We used the "average American" designation because it is used in the U.S. media and in popular culture (e.g., Corley, 2018; O'Keefe, 2012). Nonetheless, future research should adopt "average U.S. American" when U.S. citizens are recruited as participants in order to avoid lumping together different nationalities and cultures, and it may find different results.

Further, the MTurk population may differ from the general U.S. population (e.g., some found MTurk workers show lower religiosity, higher education, and lower income; Clifford et al., 2015). Therefore, when rating themselves in comparison with the "average American," MTurk workers may not be showing bias if they are rating themselves as below average in religiosity but in fact produce an accurate estimation of comparative religiosity. Future research employing MTurk samples may consider the implications of using designation of "average MTurk worker" compared to "average American" with the aim of addressing this potential confound.

Conclusion

We successfully replicated Alicke (1985). More than 30 years after the original finding, people still believe they are better than average on desirable traits. The effect of desirability on the better-than-average effect is stronger for traits considered controllable.

ORCID iD

Gilad Feldman https://orcid.org/0000-0003-2812-6599

Supplemental Material

The supplemental material is available in the online version of the article.

Note

1. We note that prior to the final preregistration of Wave 1, we had two prior preregistrations that we found had to be amended due to issues identified in the comprehension checks and a Qualtrics bug that affected randomization. Amendments were made prior to the full first data collection. Links to prior registrations (https://osf.io/a5mx7/) and (https://osf.io/pvr6t). The final preregistration for the first data collection was completed before data collection. In addition, we already preregistered the second data collection in the preregistration of the first data collection, yet following our analysis of the first data collection, we made changes to the preregistration of the section data collection. These changes are explained in the Adjustments to the Original Study subsection. For the most complete preregistration plan conducted prior to data collections, please refer to the latest preregistrations.

References

Aiken, L. S., West, S. G., & Reno, R. R. (1991). *Multiple regression: Testing and interpreting interactions*. Sage.

Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology*, *49*(6), 1621.

Alicke, M. D., & Govorun, O. (2005). The better-than-average effect. In M. D. Alicke, D. A. Dunning, & J. Krueger (Eds.), *The self in social judgment* (pp. 85–106). Psychology Press.

Alicke, M. D., Zell, E., & Guenther, C. L. (2013). Social self-analysis: Constructing, protecting, and enhancing the self. In M. Zanna & J. Olson (Eds.) *Advances in experimental social psychology* (Vol. *48*, pp. 173–234). Academic Press.

Anderson, N. H. (1968). Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology*, *9*(3), 272.

Brown, J. D. (2010). Across the (Not So) Great Divide: Cultural similarities in self-evaluative processes. *Social and Personality Psychology Compass*, *4*(5), 318–330. https://doi.org/10.1111/j.1751-9004.2010.00267.x

Brown, J. D. (2012). Understanding the better than average effect: Motives (still) matter. *Personality and Social Psychology Bulletin* *38*(2)209–219

Brown, J. D. (2014). *The self*. Psychology Press.

Chambers, J. R., & Windschitl, P. D. (2004). Biases in social comparative judgments: The role of nonmotivated factors in above-average and comparative-optimism effects. *Psychological Bulletin*, *130*(5), 813.

Chandrashekar, S. P., Yeung, S. K., Yau, K. C., Feldman, G., Cheung, C. Y., Agarwal, T. K., Wong, C. Y., Pillai, T., Thirlwell, T. N., Leung, W. N., & Li, Y. T. (2020). Agency and self-other asymmetries in perceived bias and shortcomings: Replications of the Bias Blind Spot and extensions linking to free will beliefs. Retrieved March 2020 from https://doi.org/10.13140/RG.2.2.19878.16961/1

Chen, J., Hui, L. S., Yu, T., Feldman, G., Zeng, S. V., Ching, T. L., Ng, C. H., Wu, K. W., Yuen, C. M., Lau, T. K., Cheng, B. L., & Ng, K. W. (2020, June). Foregone opportunities and choosing not to act: Replications of inaction inertia effect. *Social Psychological and Personality Science* [Manuscript accepted for publication] https://www.researchgate.net/publication/332550110_Foregone_opportunities_and_choosing_not_to_act_Replications_of_Inaction_Inertia_effect

Clifford, S., Jewell, R. M., & Waggoner, P. D. (2015). Are samples drawn from Mechanical Turk valid for research on political ideology? *Research and Politics*, *2*(4). https://doi.org/10.1177/2053168015622072

Collaborative Open-Science Research. (2020, June). Replications and extensions of classic findings in judgment and decision making. https://doi.org/10.17605/OSF.IO/5Z4A8 Retrieved June 2020 from http://osf.io/5z4a8 and http://mgto.org/pre-registered-replications/

Coppock, A. (2017). Generalizing from survey experiments conducted on Mechanical Turk: A replication approach. *Political Science Research and Methods*. https://doi.org/10.1017/psrm.2018.10

Coppock, A., Leeper, T. J., & Mullinix, K. J. (2018). Generalizability of heterogeneous treatment effect estimates across samples. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(49), 12441–12446.

Corley, T. (2018). *Average in America is a prison—Here's what it looks like and how you can break free*. CNBC.com. https://www.cnbc.com/2018/02/07/tom-corley-heres-what-average-looks-like-in-america.html

Ditto, P. H., & Griffin, J. (1993). The value of uniqueness: Self-evaluation and the perceived prevalence of valenced characteristics. *Journal of Social Behavior and Personality*, *8*(2), 221.

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, *5*(3), 69–106.

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology*, *57*(6), 1082.

Epley, N., & Dunning, D. (2000). Feeling "holier than thou": Are self-serving assessments produced by errors in self-or social prediction? *Journal of Personality and Social* Psychology, *79*(6), 861.

Frederick, S. (2012). Overestimating others' willingness to pay. *Journal of Consumer Research*, *39*(1), 1–21. https://doi.org/10.1086/662060

Guenther, C. L., Taylor, S. G., & Alicke, M. D. (2015). Differential reliance on performance outliers in athletic self-assessment. *Journal of Applied Social Psychology*, *45*(7), 374–382.

Heine, S. J., & Hamamura, T. (2007). In search of East Asian self-enhancement. *Personality and Social Psychology Review*, *11*(1), 4–27.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83; discussion 83–135. https://doi.org/10.1017/S0140525X0999152X

Irwin, J. R., & McClelland, G. H. (2003). Negative consequences of dichotomizing continuous predictor variables. *Journal of Marketing Research*, *40*(3), 366–371.

Jung, M. H., Moon, A., & Nelson, L. D. (2019). Overestimating the valuations and preferences of others. *Journal of Experimental Psychology: General*, *149*(6), 1193–1214.

Kanten, A. B., & Teigen, K. H. (2008). Better than average and better with time: Relative evaluations of self and others in the past, present, and future. *European Journal of Social Psychology*, *38*(2), 343–353.

Krizan, Z., & Suls, J. (2008). Losing sight of oneself in the above-average effect: When egocentrism, focalism, and group diffuseness collide. *Journal of Experimental Social Psychology*, *44*(4), 929–942.

Krueger, J., & Mueller, R. A. (2002). Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance. *Journal of Personality and Social Psychology*, *82*(2), 180.

LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018, April). *A unified framework to quantify the credibility of scientific findings*. http://doi.org/10.17605/OSF.IO/UWMR8

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.

Logg, J. M., Haran, U., & Moore, D. A. (2018). Is overconfidence a motivated bias? Experimental evidence. *Journal of Experimental Psychology: General*, *147*(10), 1445.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*(1), 19.

Malmendier, U., & Tate, G. (2005). CEO overconfidence and corporate investment. *The Journal of Finance*, *60*(6), 2661–2700.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, *114*(2), 376.

Mischel, W., Shoda, Y., & Ayduk, O. (2007). *Introduction to personality: Toward an integrative science of the person*. John Wiley & Sons.

Moore, D. A. (2007). Not so above average after all: When people believe they are worse than average and its implications for theories of bias in social comparison. *Organizational Behavior and Human Decision Processes*, *102*(1), 42–58.

Mullinix, K. J., Leeper, T. J., Druckman, J. N., & Freese, J. (2015). The generalizability of survey experiments. *Journal of Experimental Political Science*, *2*(2), 109–138. https://doi.org/10.1017/XPS.2015.19

O'Keefe, K. (2012). *The average American: The extraordinary search for the nation's most ordinary citizen*. Lulu Press.

Pahl, S., & Eiser, J. R. (2005). Valence, comparison focus and self-positivity biases: Does it matter whether people judge positive or negative traits? *Experimental Psychology*, *52*(4), 303–310.

Pedregon, C. A., Farley, R. L., Davis, A., Wood, J. M., & Clark, R. D. (2012). Social desirability, personality questionnaires, and the "better than average" effect. *Personality and Individual Differences*, *52*(2), 213–217.

Rothermund, K., Bak, P. M., & Brandtstädter, J. (2005). Biases in self-evaluation: Moderating effects of attribute controllability. *European Journal of Social Psychology*, *35*(2), 281–290.

Sedikides, C., Gaertner, L., & Toguchi, Y. (2003). Pancultural self-enhancement. *Journal of Personality and Social Psychology*, *84*(1), 60.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366.

Stanley, M. L., Henne, P., Iyengar, V., Sinnott-Armstrong, W., & De Brigard, F. (2017). I'm not the person I used to be: The self and autobiographical memories of immoral actions. *Journal of Experimental Psychology: General*, *146*(6), 884.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica*, *47*(2), 143–148.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, *103*(2), 193

Zell, E., Strickhouser, J. E., Sedikides, C., & Alicke, M. D. (2019). The better-than-average effect in comparative self-evaluation: A comprehensive review and meta-analysis. *Psychological Bulletin*, *146*(2), 118–149. https://doi.org/10.1037/bul0000218

Ziano, I., & Villanova, D. (2020). *More useful to you: Overestimating products' usefulness to others because of self-serving materialism attributions*. https://psyarxiv.com/938m7/

Ziano, I., Wang, Y. J., Sany, S. S., Feldman, G., Ngai, L. H., Lau, Y. K., Bhattal, I. K., Keung, P. S., Wong, Y. T., Tong, W. Z., Cheng, B. L., & Chan, H. (2020). Perceived morality of direct versus indirect harm: Replications of the preference for indirect harm effect [Manuscript accepted for publication]. https://doi.org//10.31234/osf.io/bs7jf

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, *41*, e120.

Author Biographies

Ignazio Ziano is an assistant professor at the Marketing Department, Grenoble Ecole de Management, Université Grenoble Alpes ComUE. His research focuses on consumer behavior and judgment and decision making.

Pui Yan (Cora) Mok was a guided thesis master's student working under the supervision of Gilad Feldman at the University of Hong Kong in the academic year 2018–2019.

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

Handling Editor: Will Gervais

Alicke (1985) replication and extension: Supplementary

Table of contents

Disclosures

Procedure

The replication was conducted as part of a large replication project, in which we attempted to replicate findings from the judgment and decision-making literature. In the present study, the participants from both the initial and second-wave samples received financial compensation for completing a survey.

Pre-registration

We pre-registered the study prior to data collection. The design and analysis plan were revised after our analysis of the first data collection.

All departures from the pre-registration are documented in the manuscript or in the supplementary below, see section "Pre-registration Planning and Deviation Documentation".

Data Exclusion

We pre-registered exclusion criteria such as low English proficiency and failed comprehension checks. We conducted our analyses both with and without exclusions, and found that exclusions had little effect on the results. For the sake of brevity, the manuscript reported findings without data exclusion.

Conditions Reporting

All conditions collected for this study are reported and included in the provided data.

Variables Reporting

All variables collected for this study are reported and included in the provided data.

Attention checks, Comprehension checks, and Exclusion criteria

Attention checks. 1st data collection

      To test whether participants answered the questions carefully, we added two attention checks to each condition, mixed with the trait ratings. These items were "very undesirable" and "very desirable" in the desirability condition, and "very uncontrollable" and "very controllable" in the controllability condition respectively. Participants needed to rate each of the attention check traits with the option that corresponded exactly to the trait in order to pass the attention checks. This means that participants passed the attention check if they rated "very undesirable" or "very uncontrollable" as 1 (corresponding to ""very undesirable" in the desirability condition or "very uncontrollable" in the controllability condition) and "very desirable" or "very controllable" as 7 (corresponding to "very desirable" in the desirability condition or to "very controllable" in the controllability condition).

Attention checks. 2nd data collection

      Similar to the first data collection, we added two attention checks for each condition, mixed with the trait ratings. These items were "very uncommon" and "very common" in the commonness condition, and "not at all characteristic" and "very characteristic" in the self-ratings condition respectively. Similar to the first data collection, participants were supposed to check the option "very common" for the trait "very common", very uncommon for the trait "very uncommon", "not at all characteristic" for the trait "not at all characteristic", and "very characteristic" for the trait "characteristic". Originally, we planned to use "not at all characteristic" and "very characteristic" in the other ratings condition. However, there was a coding error in the Qualtrics survey, which rendered the attention checks for this condition ineffective. This error did not impact the results reported in the following section since we pre-registered the use of a full sample for analyses. Details are reported in "Pre-registration Planning and Deviation Document" in the supplementary materials.

Comprehension questions.1st data collection.

      At the beginning of the first survey, participants received instructions about the rating criteria specific to their assigned condition: desirability or controllability. To test participants' understanding of the rating criteria, the instructions were followed by three comprehension questions with three multiple choices each. Participants had to answer all comprehension questions correctly in order to proceed to the rating task.

      In the desirability condition, participants were first asked whether a desirable characteristic or trait is one that is good to have, bad to have, or neither good or bad. Second, they were asked whether an undesirable characteristic or trait is one that is good to have, bad to have, or neither good or bad. Third, they were asked whether the task is to make evaluations based on their own desirability criteria, desirability criteria for the average American, or whatever desirability criteria seem relevant.

      In the controllability condition, participants were first asked whether a controllable characteristic or trait is one that a person could create or eliminate through sufficient effort, or a person's effort would not be sufficient to create or eliminate, or unrelated to persons. Second, they were asked whether an uncontrollable characteristic or trait is one that a person could create or eliminate through sufficient effort, or a person's effort would not be sufficient to create or eliminate, or unrelated to persons. Third, they were asked whether the task is to make evaluations based on their own controllability criteria, controllability criteria for the average American, or whatever controllability criteria seems relevant.

Comprehension questions. 2nd data collection.

At the beginning of the second survey, participants received instructions about the rating criteria specific to their assigned condition: commonness, self-ratings, or other ratings. Similar to the structure of the first survey, the instructions were followed by one to three comprehension questions with three multiple choices each. Participants had to answer all comprehension questions correctly in order to proceed to the rating task.

In the commonness condition, participants were first asked whether a common characteristic or trait is one that is frequently displayed, rarely displayed, or neither frequently nor rarely displayed. Second, they were asked the same question for an uncommon characteristic or trait. Third, they were asked whether the task is to make evaluations based on their own commonness criteria, commonness criteria for the average American, or whatever commonness criteria seems relevant.

In the self-ratings condition, they were asked whether the evaluation is based on how well the traits characterize them, the average American, or everyone. The other ratings condition comprised the same comprehension question as the self-ratings condition.

Exclusion criteria

As pre-registered, the analyses focused on the full sample. For supplementary analysis, the following exclusion criteria were pre-registered: (1) participants who reported low English proficiency (lower than 5 on a scale of 1 to 7); (2) those who reported not being serious about filling in the survey (lower than 4 on a scale of 1 to 7); (3) those who correctly guessed the study hypothesis in the funneling section; (4) those who failed to complete the survey; (5) those who failed to pass the attention check; and (6) those who completed the survey within less than one minute. Exclusion had little to no effects on results, and analyses including only participants fulfilling the preregistered criteria are reported in the "Results after Exclusions" section of these Supplementary materials.

Tables and figures

Table S1. Reported statistics and calculated effect sizes in the original study

	Reported statistics			Calculated effect sizes	
	F	df	p	η_p^2	f
Desirability	306.80	3, 261	< .0001	.78	1.88
				[.73, .81]	[1.66, 2.06]
Controllability	5.93	1, 87	< .02	.06	.26
				[.002, .18]	[0.04, 0.47]
Desirability × controllability	22.72	3, 261	< .0001	.21	.52
				[.12, .28]	[0.37, 0.62]
Desirability × controllability	14.87	1, 87	< .0005	.15	.42
				[.04, .28]	[0.2, 0.62]
Desirability × perspective	126.74	3, 261	< .0001	.59	1.21
				[.52, .65]	[1.04, 1.35]
Perspective × desirability × controllability	25.90	3, 261	< .0001	.23	.55
				[.14, .31]	[0.40, 0.66]

Note. Values in square brackets indicate the 95% confidence interval for each effect size. *df* indicates degrees of freedom. η_p^2 indicates partial eta squared. *f* indicates Cohen's *f*. Calculations can be found in "Effect Sizes and Confidence Intervals" in the supplementary materials.

Figure S1. Mean pre-ratings of traits in first-wave sample (Alicke, 1985, pp. 1629-1630).

Mean Preratings of Desirability and Controllability

Condition	Desire M	Desire rank	Control M	Control rank	Condition	Desire M	Desire rank	Control M	Control rank
High desire/high control					Moderate-high desire/high control				
Cooperative	6.28	356	5.90	328	Prideful	4.14	215	5.51	275
Considerate	6.08	341	5.96	339	Prudent	4.16	216	5.55	280
Responsible	6.28	356	5.91	334	Choosy	4.02	210	5.66	295
Friendly	6.26	355	6.05	346	M	4.32		5.77	
Respectful	5.98	324	6.36	361	Moderate-low desire/high control				
Reliable	6.37	360	5.99	341					
Resourceful	6.10	343	5.92	335	Troubled	2.01	92	6.21	355
Polite	6.16	348	6.09	349	Boastful	2.11	106	5.85	323
Dependable	6.25	353	5.90	328	Unpoised	2.29	119	5.90	328
Trustful	6.19	351	5.74	307	Jealous	2.09	103	5.89	326
Pleasant	5.98	321	5.18	288	Self-centered	2.37	123	6.05	347
Sincere	6.19	352	5.64	290	Unskilled	2.06	100	5.35	356
Loyal	5.95	320	5.69	298	Melancholy	2.31	121	5.60	285
Self-disciplined	5.94	319	5.70	301	Unsophisticated	2.45	130	5.34	287
Kind	6.08	339	5.75	308	Clumsy	2.46	131	5.66	294
Clean	5.73	292	6.37	362	Daydreamer	3.10	169	5.94	336
Good-tempered	5.88	307	5.04	328	Irreligious	2.93	160	6.02	343
Versatile	5.85	303	5.97	340	Strict	3.07	168	5.64	290
Persistent	5.47	278	5.67	297	Conforming	2.94	162	5.80	314
Well read	5.62	285	5.50	270	Compulsive	3.25	175	5.70	300
Sensitive	5.53	282	5.80	313	Hesitant	3.11	170	5.56	282
Grateful	5.53	283	5.71	304	Eccentric	3.36	179	5.77	311
Thrifty	5.38	273	5.50	270	M	2.62		5.77	
M	5.96		5.80		Low desire/high control				
Moderate-high desire/high control					Unforgiving	1.67	47	5.82	319
					Disobedient	1.87	76	6.01	342
Neat	4.83	244	5.94	337	Deceptive	1.93	84	5.94	337
Bold	5.19	260	5.84	321	Disrespectful	1.69	50	6.12	354
Self-satisfied	5.32	267	5.64	293	Snobbish	1.72	56	6.02	343
Religious	4.45	227	5.82	320	Spiteful	1.82	68	5.69	298
Self-concerned	3.58	188	6.08	348	Meddlesome	1.86	74	5.54	279
Radical	3.46	183	6.09	351	Complaining	1.91	81	5.51	301
Obedient	4.43	226	5.75	308	Unstudious	1.85	71	5.80	314
Fashionable	3.99	209	5.56	282	Uncivil	1.68	50	5.52	277

Condition	Desire M	Desire rank	Control M	Control rank
Low desire/high control				
Unappreciative	1.69	50	5.66	295
Unpleasing	1.85	72	5.55	280
Phony	1.39	5	6.11	353
Discourteous	1.58	30	5.89	326
Unkind	1.53	20	5.87	324
Rude	1.35	4	6.04	345
Impolite	1.45	11	6.10	352
Dishonest	1.27	2	6.29	359
Cold	1.56	26	5.51	275
Dishonorable	1.45	11	5.71	305
Deceitful	1.56	26	5.75	308
Hostile	1.56	27	5.61	288
Irresponsible	1.57	28	5.72	306
Unreasonable	1.62	36	5.52	277
M	1.63		5.80	
High desire/low control				
Creative	6.04	329	3.70	22
Bright	6.13	346	3.36	8
Imaginative	6.10	344	4.01	41
Intelligent	6.11	345	3.60	14
Clear-headed	6.07	336	3.46	9
Observant	6.05	330	4.04	46
Perceptive	6.08	339	4.14	58
Level-headed	6.06	334	4.10	56
Mature	6.07	336	4.35	82
Honorable	6.34	358	4.10	56
Lively	5.87	306	3.66	18
Clever	5.74	294	3.84	31
Admirable	5.74	295	3.82	29
Wise	5.85	302	3.81	28
Intellectual	5.67	287	4.06	51
Sportsmanlike	5.52	280	4.29	72
Punctual	5.45	277	4.35	82
Original	5.75	296	4.16	60
Interesting	5.83	300	4.23	67
Humerous	5.47	279	4.32	76
M	5.89		3.97	
Moderate-high desire/low control				
Reserved	3.79	196	3.59	12
Cunning	3.78	193	3.94	37
Fearless	3.84	199	4.06	53
Meticulous	4.20	217	4.34	85
Impulsive	3.79	195	4.19	81
Ordinary	3.52	186	4.39	90
Impressionable	3.48	184	4.16	60
Authoritative	3.96	206	4.36	85
Normal	4.57	233	1.90	1
Attractive	4.96	248	3.87	32
Lucky	5.02	253	3.70	21
Moderate-high desire/low control				
Ingenious	4.96	249	3.24	5
Changeable	4.61	236	1.90	1
Witty	5.20	261	3.99	41
Philosophical	4.77	241	4.04	46
Ethical	5.35	270	4.09	55
Quick	5.17	258	4.37	88
Progressive	5.17	258	4.33	79
Sharp-witted	5.33	268	4.30	73
M	4.50		3.83	
Moderate-low desire/low control				
Forgetful	2.15	114	3.76	27
Uncultured	2.41	125	3.59	12
Discontented	2.01	92	3.72	25
Dissatisfied	2.18	116	3.57	11
Withdrawn	2.07	102	3.92	36
Unoriginal	2.07	101	4.05	48
Tiresome	2.15	114	4.34	80
Profane	2.40	124	4.32	76
Unentertaining	2.33	122	4.35	82
Passive	2.71	149	3.30	7
Timid	2.56	137	3.60	14
Bashful	2.59	142	3.67	20
Restless	2.91	159	3.70	22
Unpopular	2.59	140	3.82	29
Unemotional	2.68	148	3.99	41
Meek	2.78	153	4.06	51
Overcautious	3.04	166	4.29	70
Inhibited	2.62	144	4.20	65
Extravagant	3.31	178	4.15	60
Solemn	2.56	138	4.36	85
Softspoken	3.18	173	4.22	66
M	2.54		3.95	
Low desire/low control				
Insecure	1.92	83	3.66	19
Belligerent	2.00	90	3.91	34
Humorless	1.76	48	4.05	48
Lazy	1.81	66	4.37	88
Vain	1.80	64	4.30	73
Gullible	1.96	87	4.17	87
Liar	1.24	1	3.99	40
Unpleasant	1.47	13	3.99	41
Mean	1.49	15	3.62	16
Maladjusted	1.63	39	3.56	10
Unethical	1.49	14	4.09	54
Ill-Mannered	1.55	22	4.31	75
Incompetent	1.60	33	4.27	70
Shallow	1.65	45	4.32	76
Irrational	1.63	42	4.29	70
M	1.67		4.06	

Table S2: Ratings in second-wave sample by levels of desirability and controllability in the original study

Level of control	Level of desirability			
	High	Moderate-high	Moderate-low	Low
	Ratings of self			
High	5.72 (0.57)	4.60 (0.79)	3.40 (0.73)	2.23 (0.73)
Low	5.37 (0.66)	4.60 (0.54)	3.21 (0.74)	2.59 (0.69)
	Ratings of average college student			
High	4.69 (0.72)	4.44 (0.72)	3.74 (0.61)	3.26 (0.83)
Low	4.87 (0.74)	4.27 (0.47)	3.40 (0.55)	3.40 (0.78)
	Ratings of self minus average college student			
High	1.03	0.16	-0.34	-1.03
Low	0.50	0.33	-0.19	-0.81

Note. Values in parentheses are standard deviations.

Table S3. Summary of effect sizes using commonness as the dependent variable

	Effect sizes			NHST Summary
	b	*beta*	*sr²*	
Desirability	0.16	0.53	.04	Supported
	[0.06, 0.26]	[0.21, 0.85]	[-.01, .09]	
Self-ratings	0.05	0.11	.00	Not supported
	[-0.09, 0.18]	[-0.21, 0.43]	[-.01, .01]	
Desirability × self-ratings	0.01	0.03	.00	Not supported
	[-0.03, 0.05]	[-0.10, 0.16]	[-.01, .01]	
Desirability	0.04	0.12	.01	Supported
	[0.01, 0.06]	[0.04, 0.20]	[-.00, .02]	
Other ratings	0.80	0.86	.41	Supported
	[0.72, 0.87]	[0.77, 0.94]	[.31, .52]	
Desirability × other ratings	0.02	0.03	.00	Not supported
	[-0.02, 0.07]	[-0.03, 0.10]	[-.00, .00]	

Note. *b* represents unstandardized regression weights. *beta* indicates the standardized regression weights. *sr²* represents the semi-partial correlation squared. NHST represents null hypothesis significance testing. NHST summary concerns the main effects and interactions of the following extension hypothesis: for ratings of others, trait desirability is positively associated with trait commonness. For ratings of self, trait desirability is negatively associated with trait commonness.

Table S1. Comparison of study characteristics between the original article and the replication

Study	Alicke (1985)		Replication	
Sample	Initial	Final	Initial	Final
n	80 (desirability) / 84 (controllability)	88 (self) / 88 (other)	341 (desirability) / 329 (controllability)	300 (self) / 306 (other) / 297 (commonness)
% Female	57.9	58.0	47.2	54.4
Age *M* (Years)	Unreported	Unreported	39.12	39.34
Age *SD* (Years)	Unreported	Unreported	12.01	12.42

Table S2. Summary of study design

Hypothesis 1 (Replication)		
	IV 1: Desirability	
IV 2 Condition 1: Self-perspective	**DV**: Title: Self-minus-other ratings of the traits	
IV 2 Condition 2: Other perspective	Specific DV item: Rate to which degree each trait characterizes you/the average American on a 7-point scale (1 = not at all characteristic; 7 = very characteristic).	
Hypothesis 2 (Replication)		
	IV 1: Desirability	**IV 2:** Controllability
IV 3 Condition 1: Self-perspective	**DV**: Title: Self-minus-other ratings of the traits	
IV 3 Condition 2: Other perspective	Specific DV item: Rate to which degree each trait characterizes you/the average American on a 7-point scale (1 = not at all characteristic; 7 = very characteristic).	
Hypothesis 3 (Extension)		
	IV 1: Desirability	
IV 2 Condition 1: Self-perspective	**DV**: Title: Commonness ratings of the traits	
IV 2 Condition 2: Other perspective	Specific DV item: Rate to which degree each trait is common among the average Americans on a 7-point scale (1 = not at all common; 7 = very common).	

Note. IV represents independent variable. DV represents dependent variable. In the present study, self-minus-other ratings were calculated by subtracting other ratings from other ratings to account for the rating perspective.

Table S7. Regression results using commonness as the dependent variable, desirability and self-ratings as the independent variables

Predictor	b	b 95% CI [LL, UL]	beta	beta 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	Fit	Difference
(Intercept)	4.07***	[4.00, 4.13]					R^2 = .42***	
Desirability	0.16**	[0.07, 0.26]	0.54	[0.22, 0.86]	.04	[-.01, .10]	95% CI [.29,.51] $F(2, 146) = 51.88***$	
Self-ratings	0.05	[-0.09, 0.18]	0.11	[-0.21, 0.43]	.00	[-.01, .01]		
(Intercept)	4.05***	[3.93, 4.16]						
Desirability	0.16**	[0.06, 0.26]	0.53	[0.21, 0.85]	.04	[-.01, .09]	R^2 = .42***	ΔR^2 = .001
Self-ratings	0.05	[-0.09, 0.18]	0.11	[-0.21, 0.43]	.00	[-.01, .01]	95% CI [.29, .51] $F(3, 145) = 34.48***$	95% CI [-.01, .01] $\Delta F(1, 145) = 0.22$
Interaction	0.01	[-0.03, 0.05]	0.03	[-0.10, 0.16]	.00	[-.01, .01]		

Note. A significant *b*-weight indicates the beta-weight and semi-partial correlation are also significant. *b* represents unstandardized regression weights. *beta* indicates the standardized regression weights. sr^2 represents the semi-partial correlation squared. *r* represents the zero-order correlation. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.
* indicates $p < .05$. ** indicates $p < .01$. *** indicates $p < .001$.

Table S8. Regression results using commonness as the dependent variable, desirability and other-ratings as the independent variables

Predictor	b	b 95% CI [LL, UL]	beta	beta 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	Fit	Difference
(Intercept)	4.07***	[4.03, 4.10]					R^2 = .86***	
Desirability	0.04**	[0.02, 0.06]	0.13	[0.05, 0.21]	.01	[-.00, .02]	95% CI [.82,.89]	
Other ratings	0.79***	[0.71, 0.86]	0.84	[0.77, 0.92]	.45	[.34, .56]	$F(2, 146) = 451***$	
(Intercept)	4.05***	[4.01, 4.10]					R^2 = .86***	ΔR^2 = .001
Desirability	0.04**	[0.01, 0.06]	0.12	[0.04, 0.20]	.01	[-.00, .02]	95% CI [.82,.89] $F(3, 145) = 300.8***$	95% CI [-.00, .00] $\Delta F(1, 145) = 0.92$
Other ratings	0.80***	[0.72, 0.87]	0.86	[0.77, 0.94]	.41	[.31, .52]		
Interaction	0.02	[-0.02, 0.07]	0.03	[-0.03, 0.10]	.00	[-.00, .00]		

Note. A significant *b*-weight indicates the beta-weight and semi-partial correlation are also significant. *b* represents unstandardized regression weights. *beta* indicates the standardized regression weights. sr^2 represents the semi-partial correlation squared. *r* represents the zero-order correlation. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.
* indicates $p < .05$. ** indicates $p < .01$. *** indicates $p <. 001$

Table S9. Simple main effects of desirability on self-minus-other ratings

Controllability	*b*	*p*	95% CI
One *SD* below mean***	0.29	< .001	[.21, .38]
One *SD* above mean***	0.60	< .001	[.52, .68]

Note. *** indicates $p < .001$. *b* represents unstandardized regression weights. 95% CI represents 95% confidence interval.

Table S10. Mean pre-ratings of the revised conditions

Revised Conditions	Mean Pre-rating	Mean Difference
Neutral-high D, high C	4.84	0.53
Neutral-high D, low C	5.06	0.32
Neutral-low D, high C	3.89	-0.17
Neutral-low D, low C	3.88	0.22

Note. D refers to desirability. C refers to controllability.

Comparison with the Original Article

The below tables summarize and explain the similarities and differences between the original article and replication study.

Table S11. Similarities and differences between the original article and replication study in the first-wave sample

Item	Explanation	
	Original Article	**Replication Study**
Instructions	Participants in the first-wave sample were asked them to judge to what extent the traits were desirable or controllable. Definitions of desirable and controllable were given (see Procedures in Section 3 for details).	Same instructions
Measures/ Stimulus	362 traits	149 traits (The article reported using 154 traits (Alicke, 1985, p. 1624) but the appendix listed only 149 traits.)
	Paper-and-pencil survey ● One booklet (either desirability or controllability) ● Sheets in randomized order (37 traits on each) ● Non-randomized choices	Online Qualtrics survey ● Randomized, evenly presented blocks for desirability and controllability ● 36-40 traits in total ● Added 2 attention checks for each condition ● Added 3 comprehension questions for each condition
	7-point bipolar scale (1 = not at all desirable or controllable, 7 = very characteristic of desirable or controllable)	Same scale
Procedure	Between-subjects design	Same design
	Participants rated all traits on either desirability or controllability	Participants were randomly assigned to rate 40 traits either desirability or controllability
Location	In groups (18 to 29 subjects); location unreported	Alone; online
Remuneration	Unreported	Participants received 0.5 USD for a task estimated at 4 minutes, which is commensurate with the federal minimum hourly wage of 7.25 USD.
Participant Population	Introductory psychology students at University of North Carolina at Chapel Hill, North Carolina	Americans recruited via Amazon Mechanical Turk (MTurk)

Table S12. Similarities and differences between the original article and replication study in the second-wave sample

Item	Explanation	
	Original Article	**Replication Study**
Instructions	Participants in the second-wave sample received the first booklet of traits for one perspective and were asked to rate to which degree the traits characterized them or the average college student. Then they received the second booklet and repeated the same process for the other perspective.	Participants in the second-wave sample were asked to rate to what extent the traits were characteristic of either them or the average American, or to what extent the traits are common among the average American.
Measures/ Stimulus	154 traits	149 traits (The article reported using 154 traits (Alicke, 1985, p. 1624) but the appendix listed only 149 traits.)
	Paper-and-pencil survey ● Two booklets (self & average college student) presented in counterbalanced order ● Sheets in randomized order (6 traits on each) ● Non-randomized choices	Online Qualtrics survey ● 3 randomized blocks: self-ratings, other ratings or commonness ratings ● Added 2 attention checks for each condition ● Added 3 comprehension questions for the commonness condition, and 1 comprehension question each for the self-condition and the other condition
	7-point bipolar scale (1 = not at all characteristic of me or the average college student, 7 = very characteristic of me or the average college student)	Same scale but we replaced "average college student" with "average American" to match with our target population
Procedure	Within-subjects design	Between-subjects design
	Participants rated all traits in both the self and other conditions	Participants were randomly assigned to rate 40 traits from the self or average American perspective
Location	In groups (18 to 29 subjects); location unreported	Alone; online
Remuneration	Unreported	Participants received 0.5 USD for a task estimated at 4 minutes, which is commensurate with the federal minimum hourly wage of 7.25 USD.
Participant Population	Introductory psychology students at University of North Carolina at Chapel Hill, North Carolina	Americans recruited via MTurk

Pre-registration Planning and Deviation Documentation

The below table summarizes the components where there were deviations from the pre-registration.

Table S13. Pre-registration planning and deviation documentation

Components in your preregistration	Location of preregistered decision/plan	Location of the rationale for the decision/plan (if any)	Were there deviations?*	If yes - describe details of deviation(s)	Rationale for deviation	How might the results be different if you had not deviated
Procedures	Page 12 of pre-registration	Page 12 of pre-registration	No	N/A	N/A	N/A
Power analysis	Page 13 of pre-registration	Page 13 of pre-registration	No	N/A	N/A	N/A
Exclusion rules	Page 13 of pre-registration	Page 13 of pre-registration	Minor	There was an error in Qualtrics, which rendered the attention checks ineffective in the "Other ratings" condition.	Results after exclusion in supplementary material	The size of the second-wave sample after exclusion would be slightly smaller.
Evaluation criteria	Page 16 of pre-registration	Page 16 of pre-registration	Minor	Commented on magnitude and direction only instead of using LeBel et al.'s (2018) framework	See discussion of the manuscript	N/A
Analyses	Page 17-19 of pre-registration	Page 17-20 of pre-registration	No	N/A	N/A	N/A
Presentation of statistics	Page 20 of pre-registration	Page 20 of pre-registration	Minor	Did not include a graph for the extension hypothesis	Weak to no moderating effects detected	N/A

Note. *Categories for deviations: Minor - Change probably did not affect results or interpretations; Major - Change likely affected results or interpretations.

Materials

Qualtrics Surveys

The full surveys, including the survey flow, randomization options and debrief, are available in .doc and .qsf file types on the OSF (see main manuscript for links).

Rating Criteria

Each participant was shown 40 of the 149 randomized traits (see list at the end of this section), and asked to rate these traits based on one of the five rating criteria below:

Desirability
For each of the following:
To what extent do these traits represent desirable or undesirable characteristics **for the average American**?

In this context, a desirable characteristic is one that the average American would perceive as being **good to have**, whereas an undesirable characteristic is one that the average American would perceive as being **bad to have**.
(1 = very undesirable; 7 = very desirable)

Controllability
To what extent do these traits represent controllable or uncontrollable characteristics for the average American?

A **controllable** characteristic is one that an average American **could create or eliminate through a sufficient amount of effort**, whereas an **uncontrollable** characteristic is one that an average American's **effort would not be sufficient to create or eliminate**.
(1 = very uncontrollable; 7 = very controllable)

Commonness
For each of the following:
To what extent are these traits common **among the average Americans**?

In this context, a common characteristic is one that the average American would frequently display, whereas an uncommon characteristic is one that the average American would rarely display.
(1 = very uncommon; 7 = very common)

Self Ratings
For each of the following:
To what extent do these traits characterize **you**?
(1 = not at all characteristic of me; 7 = very characteristic of me)

Other Ratings
For each of the following:
To what extent do these traits characterize **the average American**?
(1 = not at all characteristic of the average American; 7 = very characteristic of the average American)

The below is a full list of the traits used for participant ratings. We referenced the traits reported in the appendix of the original study.

List of Traits for Ratings

1.	Cooperative	51.	Unforgiving	101.	Impressionable
2.	Considerate	52.	Disobedient	102.	Authoritative
3.	Responsible	53.	Deceptive	103.	Normal
4.	Friendly	54.	Disrespectful	104.	Attractive
5.	Respectful	55.	Snobbish	105.	Lucky
6.	Reliable	56.	Spiteful	106.	Ingenious
7.	Resourceful	57.	Meddlesome	107.	Changeable
8.	Polite	58.	Complaining	108.	Witty
9.	Dependable	59.	Unstudious	109.	Philosophical
10.	Trustful	60.	Uncivil	110.	Ethical
11.	Pleasant	61.	Unappreciative	111.	Quick
12.	Sincere	62.	Unpleasing	112.	Progressive
13.	Loyal	63.	Phony	113.	Sharp-witted
14.	Self-disciplined	64.	Discourteous	114.	Forgetful
15.	Kind	65.	Unkind	115.	Uncultured
16.	Clean	66.	Rude	116.	Discontented
17.	Good-tempered	67.	Impolite	117.	Dissatisfied
18.	Versatile	68.	Dishonest	118.	Withdrawn
19.	Persistent	69.	Cold	119.	Unoriginal
20.	Well read	70.	Dishonorable	120.	Tiresome
21.	Sensitive	71.	Deceitful	121.	Profane
22.	Grateful	72.	Hostile	122.	Unentertaining
23.	Thrifty	73.	Irresponsible	123.	Passive
24.	Neat	74.	Unreasonable	124.	Timid
25.	Bold	75.	Creative	125.	Bashful
26.	Self-satisfied	76.	Bright	126.	Restless
27.	Religious	77.	Imaginative	127.	Unpopular
28.	Self-concerned	78.	Intelligent	128.	Unemotional
29.	Radical	79.	Clear-headed	129.	Meek
30.	Obedient	80.	Observant	130.	Overcautious
31.	Fashionable	81.	Perceptive	131.	Inhibited
32.	Prideful	82.	Level-headed	132.	Extravagant
33.	Prudent	83.	Mature	133.	Solemn
34.	Choosy	84.	Honorable	134.	Softspoken
35.	Troubled	85.	Lively	135.	Insecure
36.	Boastful	86.	Clever	136.	Belligerent
37.	Unpoised	87.	Admirable	137.	Humorless
38.	Jealous	88.	Wise	138.	Lazy
39.	Self-centered	89.	Intellectual	139.	Vain
40.	Unskilled	90.	Sportsmanlike	140.	Gullible
41.	Melancholy	91.	Punctual	141.	Liar
42.	Unsophisticated	92.	Original	142.	Unpleasant
43.	Clumsy	93.	Interesting	143.	Mean
44.	Daydreamer	94.	Humorous	144.	Maladjusted
45.	Irreligious	95.	Reserved	145.	Unethical
46.	Strict	96.	Cunning	146.	Ill-mannered
47.	Conforming	97.	Fearless	147.	Incompetent
48.	Compulsive	98.	Meticulous	148.	Shallow
49.	Hesitant	99.	Impulsive	149.	Irrational
50.	Eccentric	100.	Ordinary		

Effect Sizes and Confidence Intervals

Confidence intervals for eta-squared in the original article were calculated using the below software:
- η_p^2 calculation: https://effect-size-calculator.herokuapp.com/#partial-eta-squared-fixed-effects
- η_p^2 to f conversion: https://www.psychometrica.de/effect_size.html#transform

For effect size conversions to f, we used eta-squared to six or seven decimal places (as shown in the screenshots below) for the estimate and the values within the 95% confidence interval. In the final manuscript, we reported the f values for the effect sizes.

Main effects:
(1) Desirability:
- Reported: $F(3, 261) = 306.80, p < .0001$
- Calculated effect sizes:
  - $\eta_p^2 = .78$, 95% CI [.73, .81]
  - $f = 1.88$, 95% CI [1.66, 2.06]

Partial eta-squared (Fixed effects)

Inputs

F-value: 306.8   Confidence Interval: 95   %

Numerator degrees of freedom:
3

Denominator degrees of freedom:
261

[Calculate]   Clear

It is recommended that you use the **90% CI** if you have an alpha level of 5%.

Entered values: {
  ":f": 306.8,
  ":df_effect": 3,
  ":df_error": 261,
  ":conf_int": 95
}

Results (CI using noncentral F distribution)

Partial eta-squared: 0.7790757

Lower limit on partial eta-squared:
0.7338855

Partial omega-squared:
0.7758796

Upper limit on partial eta-squared:
0.8094648

13. Transformation of the effect sizes d, r, f, Odds Ratio and η^2

Please choose the effect size, you want to transform, in the drop-down menu. Specify the magnitude of the effect size in the text field on the right side of the drop-down menu afterwards. The transformation is done according to Cohen (1988), Rosenthal (1994, S. 239) and Borenstein, Hedges, Higgins, and Rothstein (2009; transformation of d in Odds Ratios).

Effect Size	Eta Square	0.7790757
d		3.7558
r		0.8827
η^2		0.7791
f		1.8779
Odds Ratio		908.8672
Number Needed to Treat (NNT)		1.008

(2) Controllability:
- Reported: $F(1, 87) = 5.93$, $p < .02$
- Calculated:
  - $\eta_p^2 = .06$, 95% CI [.002, .18]
  - $f = .26$, 95% CI [0.04, 0.47]

Partial eta-squared (Fixed effects)

Inputs

F-value: 5.93   Confidence Interval: 95 %

Numerator degrees of freedom: 1   Denominator degrees of freedom: 87

[Calculate]   [Clear]

It is recommended that you use the **90% CI** if you have an alpha level of 5%.

Entered values: {
":f": 5.93,
":df_effect": 1,
":df_error": 87,
":conf_int": 95
}

Results (CI using noncentral F distribution)

Partial eta-squared: 0.0638115   Lower limit on partial eta-squared: 0.001669

Partial omega-squared: 0.0524859   Upper limit on partial eta-squared: 0.1800908

13. Transformation of the effect sizes *d, r, f, Odds Ratio* and η^2

Please choose the effect size, you want to transform, in the drop-down menu. Specify the magnitude of the effect size in the text field on the right side of the drop-down menu afterwards. The transformation is done according to Cohen (1988), Rosenthal (1994, S. 239) and Borenstein, Hedges, Higgins, and Rothstein (2009; transformation of d in Odds Ratios).

Effect Size	Eta Square	0.0638115
d	0.5222	
r	0.2526	
η^2	0.0638	
f	0.2611	
Odds Ratio	2.5782	
Number Needed to Treat (NNT)	3.4718	

Interactions:

(1) Desirability x controllability (for revised categorisation: high, neutral-high, neutral-low, low desirability):

- Reported: $F(1, 87) = 14.87$, $p < .0005$
- Calculated:
  - $\eta_\text{p}^2 = 0.15$, 95% CI [.04, .28]
  - $f = .42$, 95% CI [0.19, 0.62]

Partial eta-squared (Fixed effects)

Inputs

F-value: 14.87   Confidence Interval: 95   %

Numerator degrees of freedom: 1   Denominator degrees of freedom: 87

[Calculate]   Clear

It is recommended that you use the **90% CI** if you have an alpha level of 5%.

Entered values: {
"f": 14.87,
":df_effect": 1,
":df_error": 87,
":conf_int": 95
}

Results (CI using noncentral F distribution)

Partial eta-squared: 0.1459704   Lower limit on partial eta-squared: 0.0352826

Partial omega-squared: 0.1348304   Upper limit on partial eta-squared: 0.2803132

13. Transformation of the effect sizes d, r, f, Odds Ratio and η^2

Please choose the effect size, you want to transform, in the drop-down menu. Specify the magnitude of the effect size in the text field on the right side of the drop-down menu afterwards. The transformation is done according to Cohen (1988), Rosenthal (1994, S. 239) and Borenstein, Hedges, Higgins, and Rothstein (2009; transformation of d in Odds Ratios).

Effect Size	Eta Square	0.1459704
d		0.8268
r		0.3821
η^2		0.146
f		0.4134
Odds Ratio		4.4805
Number Needed to Treat (NNT)		2.2664

(2) Desirability x controllability:
- Reported: $F(3, 261) = 22.72$, $p < .0001$
- Calculated:
  - $\eta_p^2 = 0.21$, 95% CI [.12, .28]
  - $f = .51$, 95% CI [0.37, 0.63]

Partial eta-squared (Fixed effects)

Inputs

F-value: 22.72   Confidence Interval: 95   %

Numerator degrees of freedom: Denominator degrees of freedom:
3   261

[Calculate]   [Clear]

It is recommended that you use the **90% CI** if you have an alpha level of 5%.

Entered values: {
  ":f": 22.72,
  ":df_effect": 3,
  ":df_error": 261,
  ":conf_int": 95
}

Results (CI using noncentral F distribution)

Partial eta-squared: 0.2070725   Lower limit on partial eta-squared:
                                                  0.120336

Partial omega-squared: Upper limit on partial eta-squared:
0.1973589   0.2829254

13. Transformation of the effect sizes d, r, f, Odds Ratio and η^2

Please choose the effect size, you want to transform, in the drop-down menu. Specify the magnitude of the effect size in the text field on the right side of the drop-down menu afterwards. The transformation is done according to Cohen (1988), Rosenthal (1994, S. 239) and Borenstein, Hedges, Higgins, and Rothstein (2009; transformation of d in Odds Ratios).

Effect Size	Eta Square	0.2070725
d		1.0221
r		0.4551
η^2		0.2071
f		0.511
Odds Ratio		6.3841
Number Needed to Treat (NNT)		1.8863

(2) Desirability x perspective:
- Reported: $F(3, 261) = 126.74$, $p < .0001$
- Calculated:
    - $\eta_p^2 = 0.59$, 95% CI [.52, .65]
    - $f = 1.21$, 95% CI [1.04, 1.35]

Partial eta-squared (Fixed effects)

Inputs

F-value: 126.74   Confidence Interval: 95   %

Numerator degrees of freedom: Denominator degrees of freedom:
3   261

Calculate   Clear

It is recommended that you use the **90% CI** if you have an alpha level of 5%.

Entered values: {
    ":f": 126.74,
    ":df_effect": 3,
    ":df_error": 261,
    ":conf_int": 95
}

Results (CI using noncentral F distribution)

Partial eta-squared: 0.5929634   Lower limit on partial eta-squared:
                                         0.5173856

Partial omega-squared: Upper limit on partial eta-squared:
0.5873688   0.6465524

13. Transformation of the effect sizes d, r, f, Odds Ratio and η^2

Please choose the effect size, you want to transform, in the drop-down menu. Specify the magnitude of the effect size in the text field on the right side of the drop-down menu afterwards. The transformation is done according to Cohen (1988), Rosenthal (1994, S. 239) and Borenstein, Hedges, Higgins, and Rothstein (2009; transformation of d in Odds Ratios).

Effect Size	Eta Square	0.5929634
d		2.4139
r		0.77
η^2		0.593
f		1.207
Odds Ratio		79.7112
Number Needed to Treat (NNT)		1.0963

(3) Perspective x desirability x controllability:
- Reported: $F(3, 261) = 25.90$, $p < .0001$
- Calculated:
  - $\eta_p^2 = .23$, 95% CI [.14, .31]
  - $f = .55$, 95% CI [.40, .66]

Partial eta-squared (Fixed effects)

Inputs

F-value: 25.90   Confidence Interval: 95   %

Numerator degrees of freedom: Denominator degrees of freedom:
3   261

[Calculate]   [Clear]

It is recommended that you use the **90% CI** if you have an alpha level of 5%.

Entered values: {
  ":f": 25.9,
  ":df_effect": 3,
  ":df_error": 261,
  ":conf_int": 95
}

Results (CI using noncentral F distribution)

Partial eta-squared: 0.2294066   Lower limit on partial eta-squared:
                                                  0.140186

Partial omega-squared: Upper limit on partial eta-squared:
0.2198999   0.3059019

13. Transformation of the effect sizes d, r, f, Odds Ratio and η^2

Please choose the effect size, you want to transform, in the drop-down menu. Specify the magnitude of the effect size in the text field on the right side of the drop-down menu afterwards. The transformation is done according to Cohen (1988), Rosenthal (1994, S. 239) and Borenstein, Hedges, Higgins, and Rothstein (2009; transformation of d in Odds Ratios).

Effect Size	Eta Square	0.2294066
d		1.0912
r		0.479
η^2		0.2294
f		0.5456
Odds Ratio		7.2376
Number Needed to Treat (NNT)		1.7868

Power Analyses

Using G*Power Version 3.1.9.3, we conducted the below power analysis to derive a minimum sample size of 71 participants. Below is the protocol of the power analysis:

F tests - ANOVA: Fixed effects, special, main effects and interactions

Analysis: A priori: Compute required sample size
Input: Effect size f   =   0.511
                   α err prob   =   0.05
                   Power (1-β err prob)   =   0.95
                   Numerator df   =   3
                   Number of groups   =   8
Output: Noncentrality parameter λ =   18.5395910
                   Critical F   =   2.7505411
                   Denominator df   =   63
                   Total sample size   =   71
                   Actual power   =   0.9528557

Note. We pasted the incorrect power analysis protocol in an earlier version of a pre-registration, using ANCOVA. However, it did not affect the final sample size.

Statistical Assumptions and Normality Tests

We conducted a series of tests of statistical assumptions for analyses. These tests include: a) residual analysis (using residuals versus fitted plot) and normality of residuals (using Q-Q plot). Below are the plots for the results before and after exclusion.

Before Exclusion

Figure S2. Residuals versus fitted plot for self-minus-other ratings predicted from desirability and controllability before exclusion.
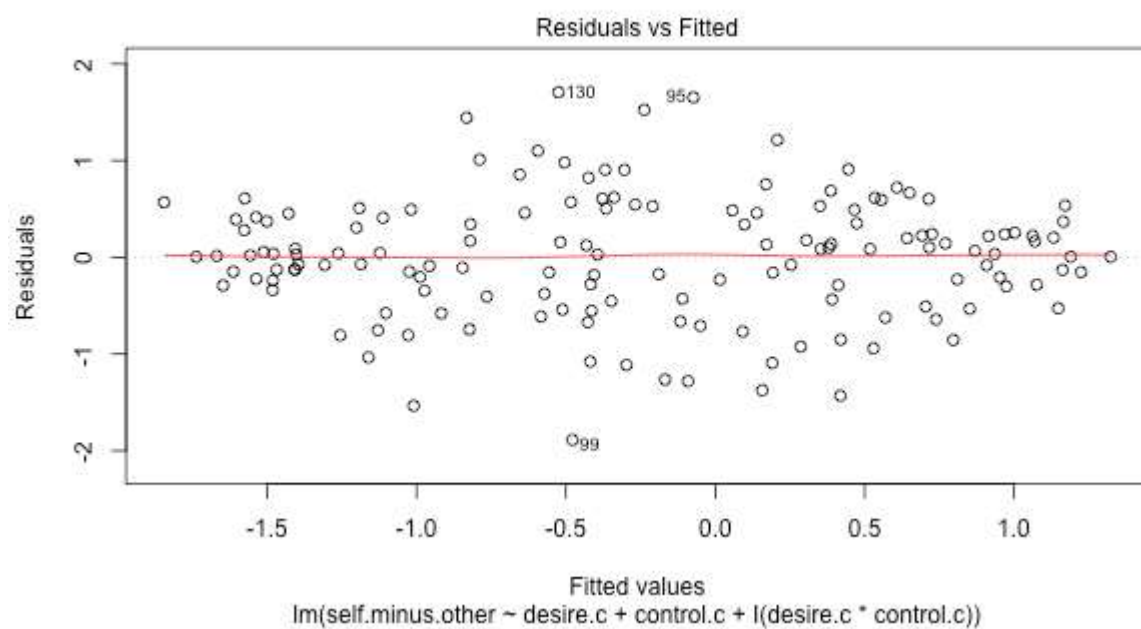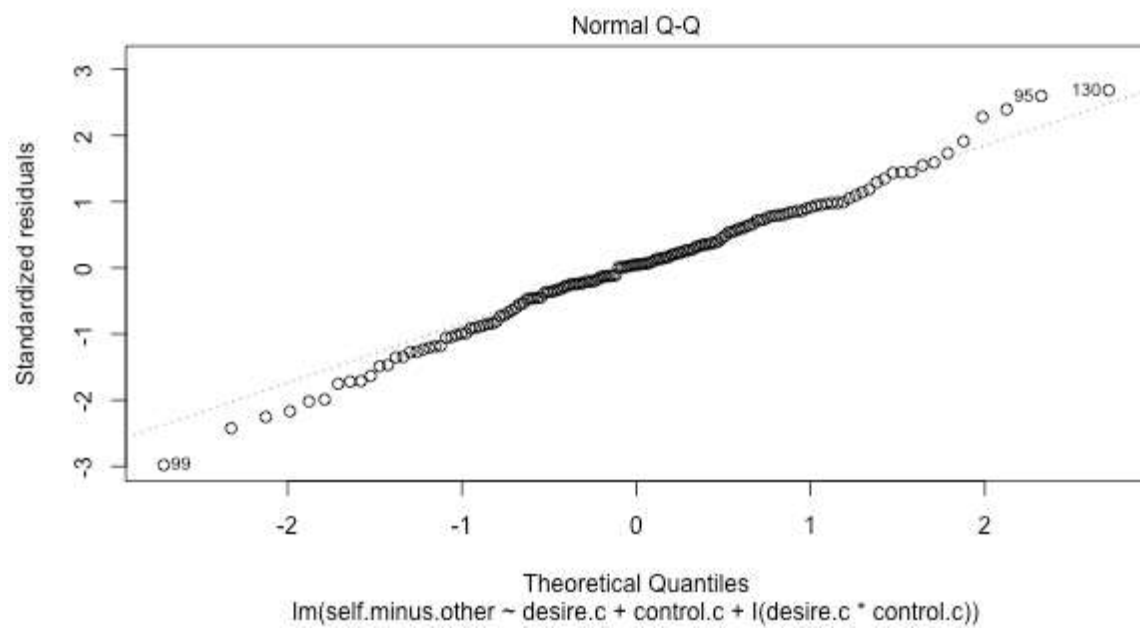
Figure S3. Normal Q-Q plot for self-minus-other ratings predicted from desirability and controllability before exclusion.

Figure S4. Residuals versus fitted plot for commonness predicted from desirability and self-ratings before exclusion

.
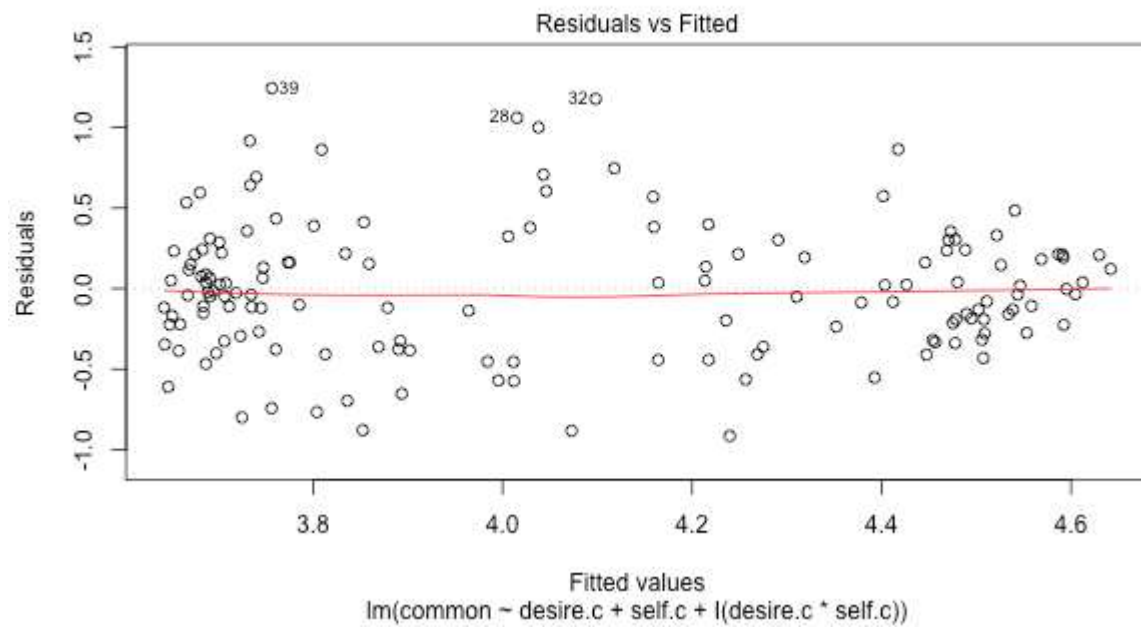
Figure S5. Normal Q-Q plot for commonness predicted from desirability and self-ratings before exclusion.
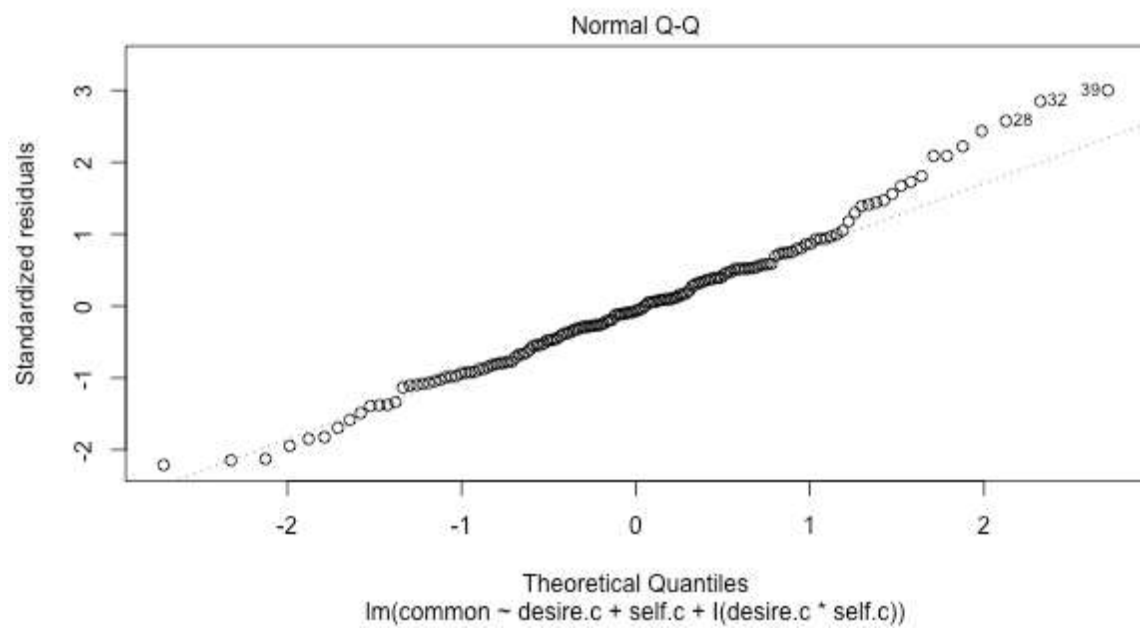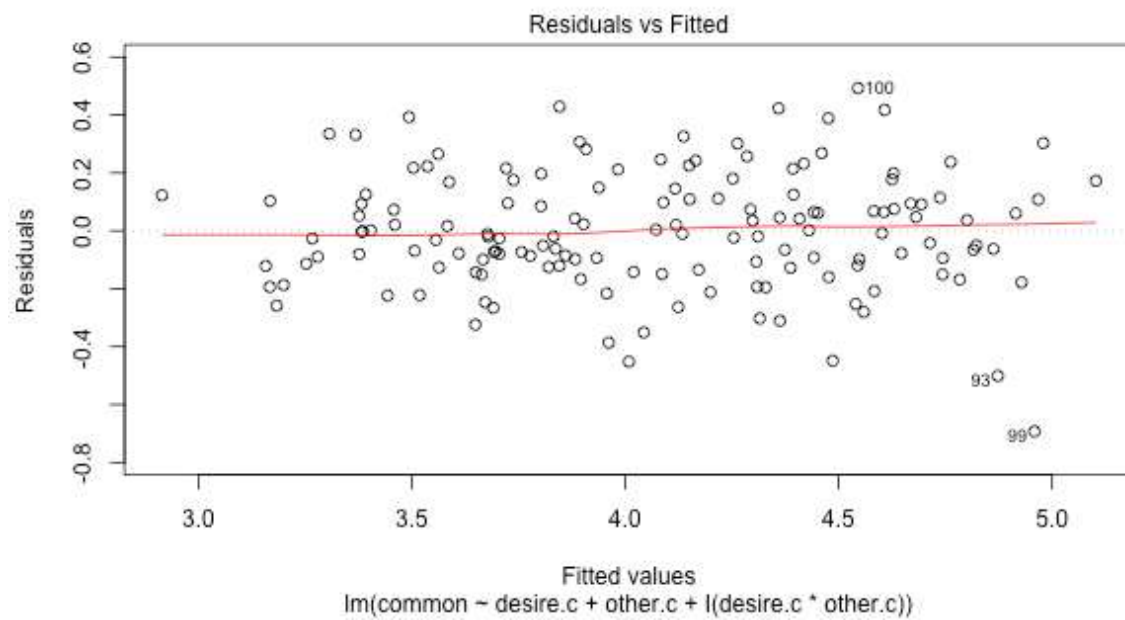
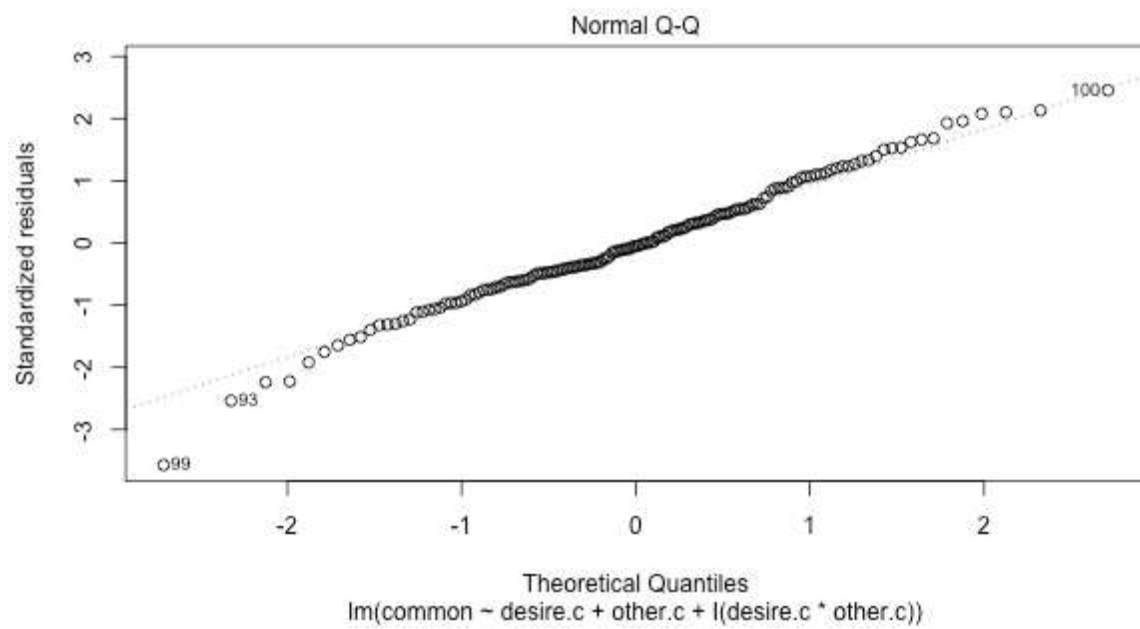Figure S6. Residuals versus fitted plot for commonness predicted from desirability and other ratings before exclusion.

Figure S7. Normal Q-Q plot for commonness predicted from desirability and other ratings before exclusion.

After Exclusion

Figure S8. Residuals versus fitted plot for self-minus-other ratings predicted from desirability and controllability after exclusion.
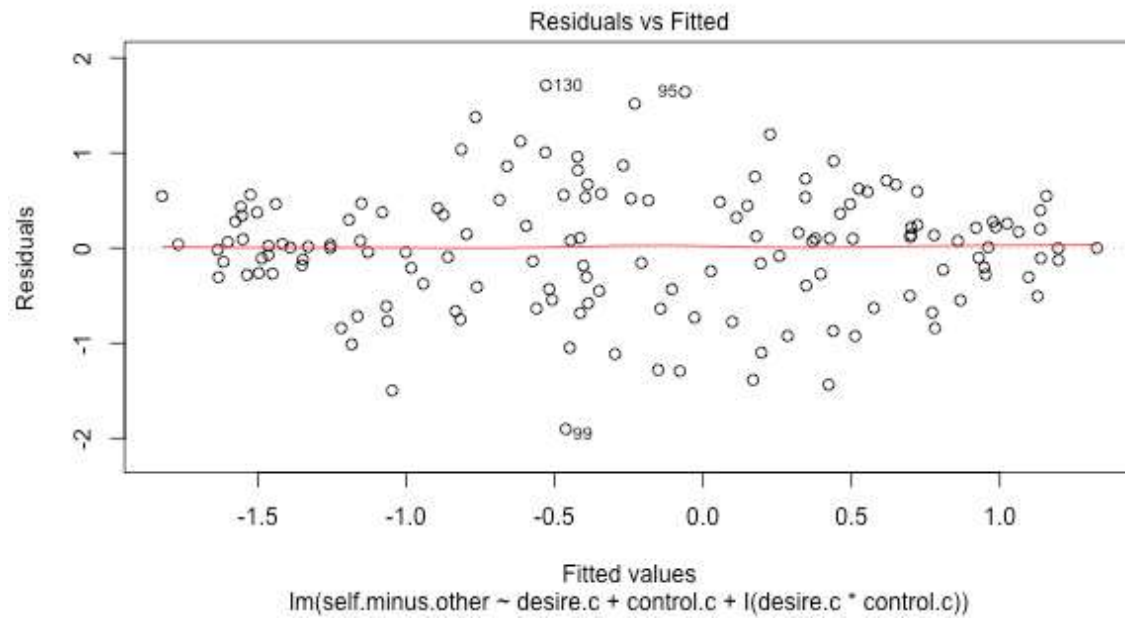
Figure S9. Normal Q-Q plot for self-minus-other ratings predicted from desirability and controllability after exclusion
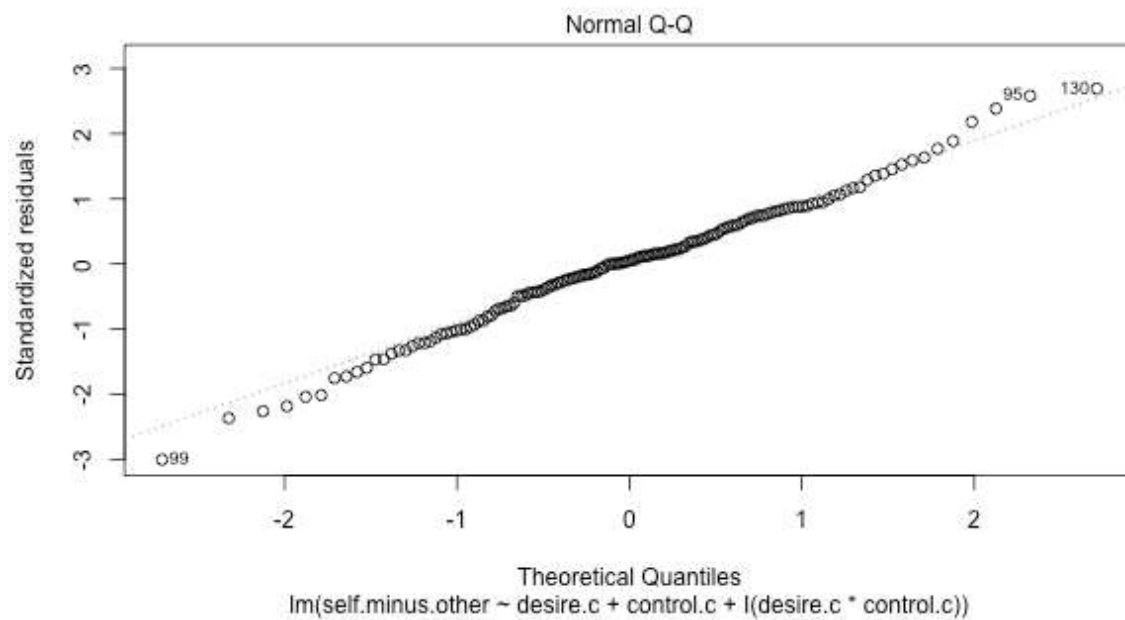
Figure S10. Residuals versus fitted plot for commonness predicted from desirability and self-ratings after exclusion.
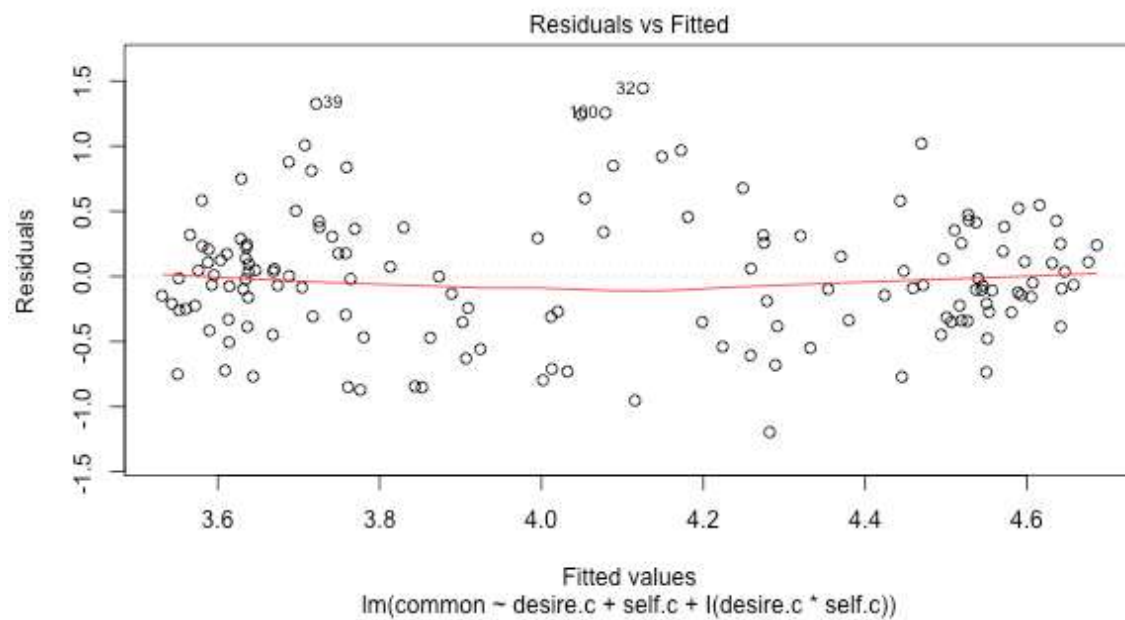
Figure S11. Normal Q-Q plot for commonness predicted from desirability and self-ratings after exclusion.
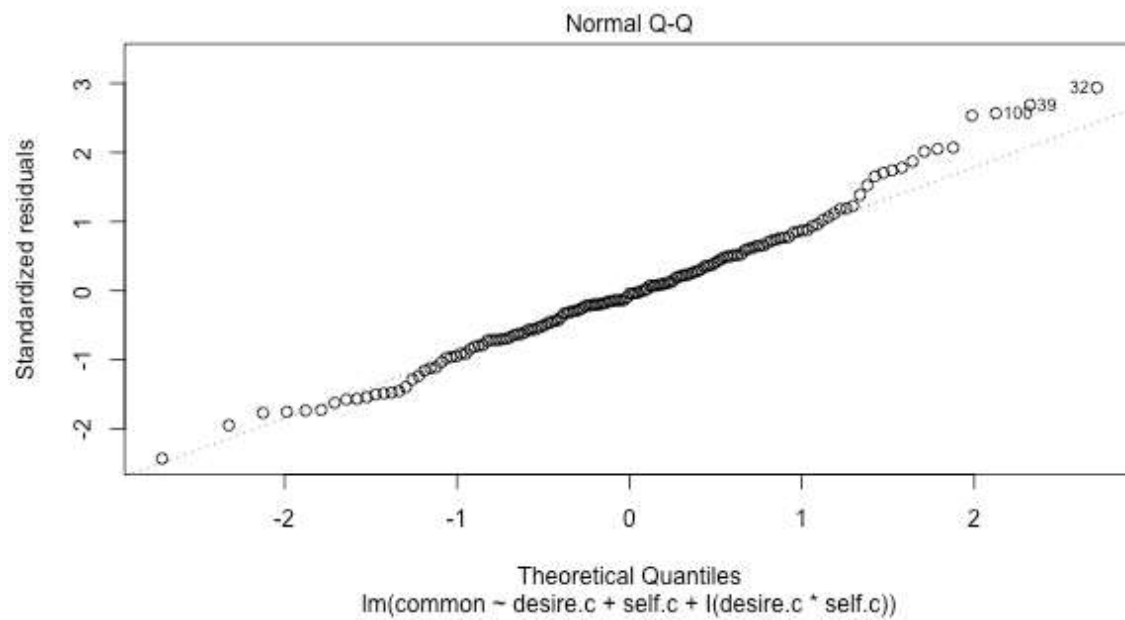
Figure S12. Residuals versus fitted plot for commonness predicted from desirability and other ratings after exclusion.
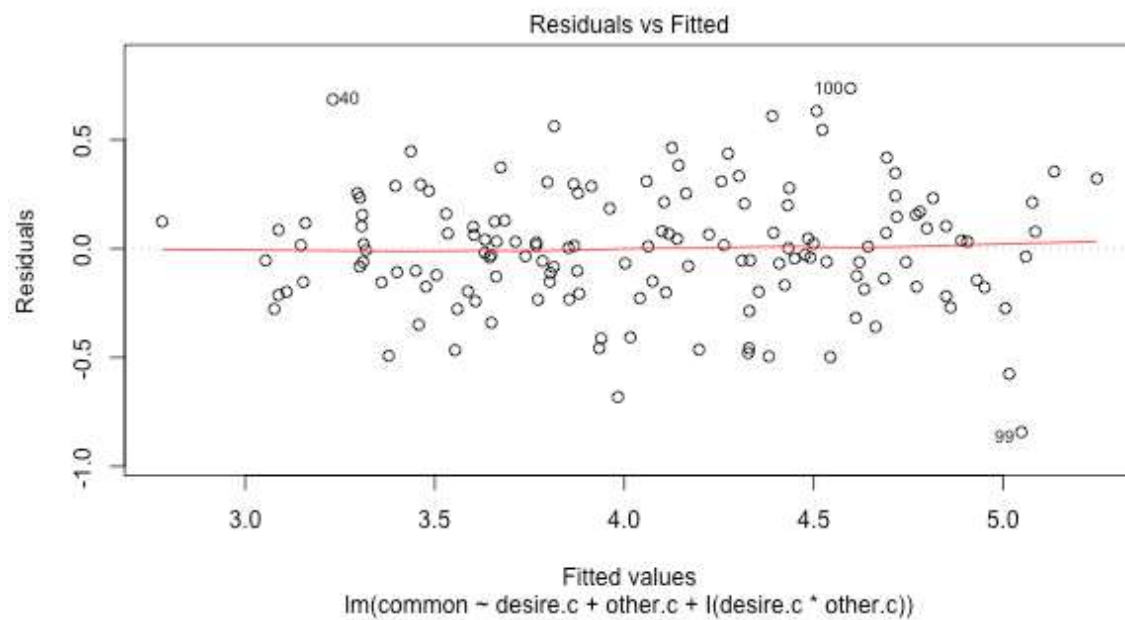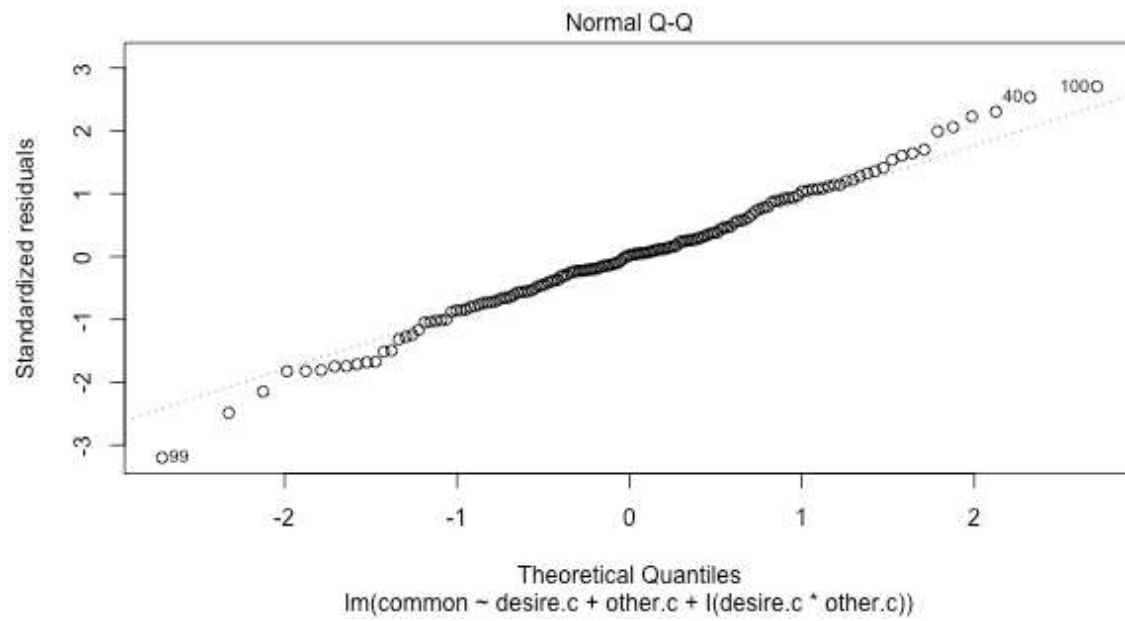
Figure S13. Normal Q-Q plot for commonness predicted from desirability and other ratings after exclusion.

Normal Q-Q

$lm(common \sim desire.c + other.c + I(desire.c * other.c))$

Exploratory Analyses

The correlation comparison between desirability and self, and desirability and other was pre-registered as one of the main analyses, whereas the remaining two correlation correlations were beyond our pre-registration. We report the results and the effect sizes of all three comparisons in this section, so that readers can compare the strengths of these differences.

We compared correlations between the study variables, using the R package cocor (Diedenhofen & Musch, 2015). Since the package is limited to comparisons of only two correlations, we focused on only the main effects for the hypotheses involving more than one predictor. Comparisons were based on dependent groups with overlapping variables. Using the results, effect sizes were computed using Lakens' (2013) spreadsheet calculator. Results are summarized in Table 19.

Table S14. Summary of correlation comparisons and effect sizes

Correlations	$r.\text{jk} - r.\text{jh}$	t	p	q
Desire, self | Desire, other***	.31 [.22, .42]	10.67	<.001 (one-tailed)	0.88
Self-minus-other, desire | Self-minus-other, control***	.79 [.62, .96]	10.72	<.001 (one-tailed)	1.04
Common, desire | Common, self-minus-other***	.42 [.32, .53]	11.34	<.001	0.76

Observations: 149

Note. $r.\text{jk} - r.\text{jh}$ refers to the difference between the correlations. $r.\text{kh}$ refers to the related correlation. *** indicates $p < .001$. Hendrickson, Stanley, and Hills' (1970) t values are reported. Hittner, May and Silver's (2003) z values are reported. q indicates Cohen's q, an effect size used for measuring correlational difference.

Results after Exclusion

The below tables summarize the results after excluding data that met our pre-registered criteria. For details of the criteria, please refer to the Replication Recipe.

The full exclusion criteria apply to all conditions, except the "Other ratings" condition in the second-wave sample. For this condition, we removed failure to pass attention checks from the criteria given an error in the Qualtrics survey. For the attention checks of this condition, "very common" and "not at all common" were used instead of "very characteristic of the average American" and "not at all characteristic of the average American". Since this error undermined the attention checks, participants were only excluded if they met the other exclusion criteria, such as English proficiency and seriousness towards the survey.

Table S15. Summary of demographics of the first-wave and second-wave samples after exclusion

	First-wave ($n = 607$)	Second-wave ($n = 771$)
Gender		
Male	309 (50.9%)	346 (44.9%)
Female	294 (48.4%)	423 (54.9%)
Missing	4 (0.7%)	2 (0.3%)
Age		
Mean (SD)	39.3 (12.1)	39.5 (12.4)
Median [Min, Max]	36.0 [18.0, 77.0]	37.0 [18.0, 87.0]
Missing	4 (0.7%)	2 (0.3%)

Table S16. Means, standard deviations, and correlations with confidence intervals after exclusion

Variable	M	SD	Desirability	Controllability	Commonness	Self-ratings	Other-ratings
Desirability	3.73	1.78					
Controllability	4.94	0.91	.01 [-.15, .17]				
Commonness	4.07	0.54	.64** [.54, .73]	.21* [.05, .36]			
Self-ratings	3.73	1.28	.92** [.89, .94]	.02 [-.14, .18]	.61** [.50, .70]		
Other-ratings	4.23	0.85	.02 [-.14, .18]	.06 [-.10, .22]	.52** [.39, .63]	.01 [-.15, .17]	
Self-minus-other ratings	-0.50	1.53	.76** [.68, .82]	-.02 [-.18, .14]	.22** [.06, .37]	.83** [.77, .88]	-.55** [-.65, -.42]

Note. M and SD are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014). * indicates $p < .05$. ** indicates $p < .01$.

Figure S14. Scatterplot showing the relationship between desirability and self-minus-other ratings with 95% confidence interval after exclusion.
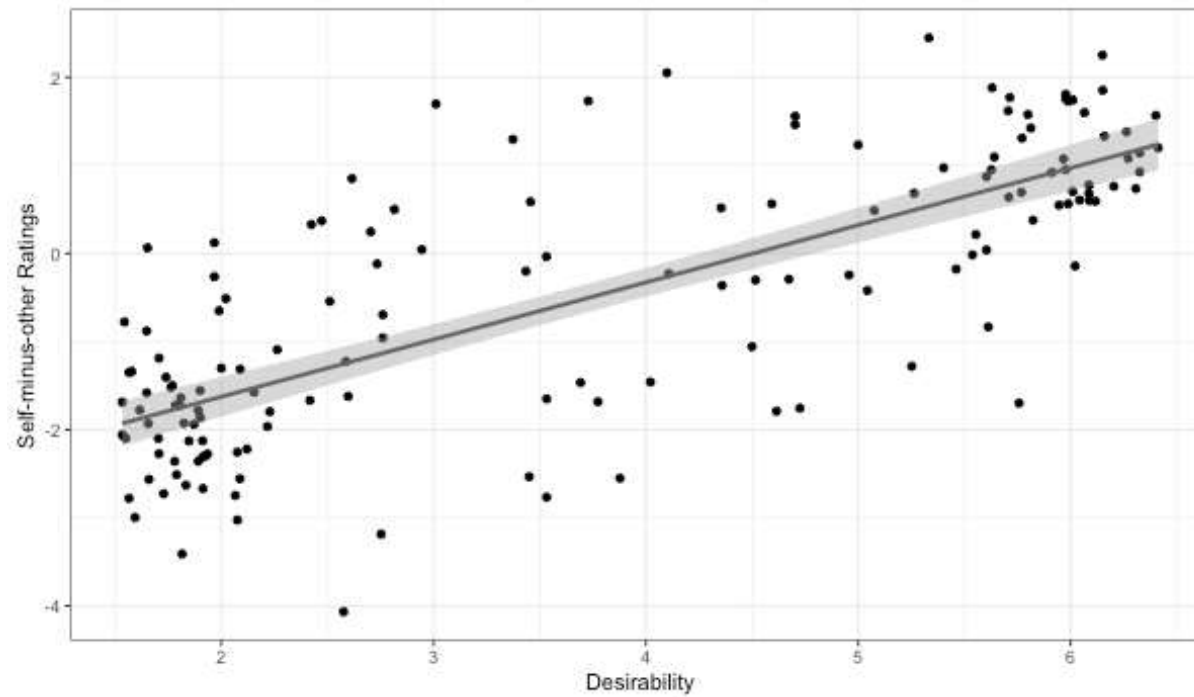
Table S17

Regression results using self-minus-other ratings as the criterion after exclusion

Predictor	b	b 95% CI [LL, UL]	beta	beta 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	Fit	Difference
(Intercept)	-0.24***	[-0.36, -0.13]					R^2 = .59***	
Desirability	0.46***	[0.40, 0.53]	0.77	[0.66, 0.87]	.59	[.48, .69]	95% CI [.48,.66]	
Controllability	-0.07	[-0.19, 0.05]	-0.06	[-0.16, 0.04]	.00	[-.01, .02]	F(2, 146) = 104.7***	
(Intercept)	-0.25***	[-0.35, -0.14]					R^2 = .66***	ΔR^2 = .07***
Desirability	0.45***	[0.39, 0.51]	0.74	[0.64, 0.84]	.54	[.43, .65]	95% CI [.56,.71] F(3, 145) = 91.9***	95% CI [.02, .12] F(1, 145) = 27.88***
Controllability	-0.12*	[-0.24, -0.01]	-0.11	[-0.20, -0.01]	.01	[-.01, .03]		
Interaction	0.17***	[0.10, 0.23]	0.26	[0.16, 0.36]	.07	[.02, .12]		

Note. A significant *b*-weight indicates the beta-weight and semi-partial correlation are also significant. *b* represents unstandardized regression weights. *beta* indicates the standardized regression weights. sr^2 represents the semi-partial correlation squared. *r* represents the zero-order correlation. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.
* indicates $p < .05$. ** indicates $p < .01$. *** indicates $p < .001$.

Table S18. Regression results using commonness as the criterion after exclusion

Predictor	b	b 95% CI [LL, UL]	beta	beta 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	Fit	Difference
(Intercept)	4.07***	[3.99, 4.15]					$R^2 = .39$***	
Desirability	0.17**	[0.05, 0.28]	0.47	[0.14, 0.79]	.03	[-.01, .08]	95% CI [.27, .49]	
Self-ratings	0.08	[-0.08, 0.24]	0.17	[-0.15, 0.50]	.00	[-.01, .02]	$F(2, 146) = 47.35$***	
(Intercept)	4.07***	[3.93, 4.20]					$R^2 = .39$***	$\Delta R^2 = .000$
							95% CI [.26, .49]	95% CI [-.00, .00]
Desirability	0.17**	[0.05, 0.28]	0.47	[0.14, 0.80]	.03	[-.01, .08]	$F(3, 145) = 31.35$***	$F(1, 145) = 0.002$
Self-ratings	0.08	[-0.08, 0.24]	0.17	[-0.16, 0.50]	.00	[-.01, .02]		
Interaction	-0.00	[-0.05, 0.05]	-0.00	[-0.13, 0.13]	.00	[-.00, .00]		

Note. A significant b-weight indicates the beta-weight and semi-partial correlation are also significant. b represents unstandardized regression weights. *beta* indicates the standardized regression weights. sr^2 represents the semi-partial correlation squared. r represents the zero-order correlation. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.
* indicates $p < .05$. ** indicates $p < .01$. *** indicates $p < .001$.

Table S19. Regression results using commonness as the criterion after exclusion

Predictor	b	b 95% CI [LL, UL]	beta	beta 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	Fit	Difference
(Intercept)	4.07***	[4.02, 4.11]					R^2 = .81**	
Desirability	0.04**	[0.01, 0.08]	0.13	[0.04, 0.22]	.01	[-.00, .02]	95% CI [.75,.84] $F(2, 146) = 309.7***$	
Other ratings	0.89***	[0.79, 0.99]	0.82	[0.73, 0.91]	.42	[.31, .53]		
(Intercept)	4.03***	[3.97, 4.09]					R^2 = .81**	ΔR^2 = .003
Desirability	0.04*	[0.00, 0.07]	0.10	[0.01, 0.20]	.01	[-.00, .02]	95% CI [.76,.85] $F(3, 145) = 209.5***$	95% CI [-.00, .01] $F(1, 145) = 2.53$
Other ratings	0.92***	[0.81, 1.02]	0.84	[0.75, 0.94]	.40	[.29, .51]		
Interaction	0.05	[-0.01, 0.11]	0.06	[-0.01, 0.14]	.00	[-.00, .01]		

Note. A significant *b*-weight indicates the beta-weight and semi-partial correlation are also significant. *b* represents unstandardized regression weights. *beta* indicates the standardized regression weights. sr^2 represents the semi-partial correlation squared. *r* represents the zero-order correlation. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.
* indicates $p < .05$. ** indicates $p < .01$. *** indicates $p < .001$