PERSPECTIVE/OPINION

Self-enhancement: Conceptualization and Assessment

Joachim I. Krueger*, Patrick R. Heck† and Jens B. Asendorpf‡

Self-enhancement bias is conventionally construed as an unwarranted social comparison in social psychology and a misperception of social reality in personality psychology. Researchers in both fields rely heavily on discrepancy scores to represent self-enhancement and fail to distinguish between a general tendency or bias to self-enhance and a self-enhancement error, or *false* perception of own excellence. We critically review prominent discrepancy measures and then propose a decision-theoretic alternative that [a] mitigates confounds between self-positivity and self-superiority, [b] separates error from bias, and [c] discourages reliance on measures that reify self-enhancement as a stable personality trait. To evaluate our hypotheses, we re-analyze data collected in our laboratory and perform a series of simulation studies. We share these materials with interested researchers.

"The only thing that sustains one through life is the consciousness of the immense inferiority of everybody else, and this is a feeling that I have always cultivated."

— Oscar Wilde (1888/2013)

"A human ego, like a gas, will always expand unless restrained by external pressure."

— Bertrand Russell (1926/1994)

Oscar Wilde's *Remarkable Rocket* flaunted his penchant for self-enhancement and his eagerness to preserve it. Bertrand Russell took a dimmer view, hinting at the vacuity of self-enhancement and it being commonplace rather than élite. Psychological research corroborates aspects of both Wilde's and Russell's propositions. Consistent with Wilde, self-enhancement is at least partly the outcome of motivated reasoning; consistent with Russell, self-enhancement is widespread (Alicke & Sedikides, 2011). In this sense, the Remarkable Rocket was not so remarkable after all. The scientific study of self-enhancement falls at the intersection of personality, self-perception, and social judgment. Across these fields, most researchers agree on the basic conceptualization self-enhancement as the perception of the self in overly positive terms. The qualifier "overly" is critical. It signals an unwarranted inflation of the self and suggests the availability of an objective standard by which this inflation can be judged. Researchers

show less agreement on how to think about this inflation and how to study it. Theories and measurement techniques are diverse, rendering the area a fuzzy set with family resemblance rather than a coherent category with an essential core.

In this article, we review some of the most prominent measurement methods of self-enhancement in light of the theories that gave rise to them. We begin with a focus on a family of discrepancy-score indices. Our assumption is that many researchers are not aware of subtle yet critical differences among these indices. Occasionally, we see investigators switching from one discrepancy-score to another without comment. Another assumption is that discrepancy scores facilitate the *reification* of measurement. Personality traits are often deliberately reified to be treated as latent constructs residing in the mind-brain. These constructs may not be directly observable, but they can be inferred from response patterns. Measurement scales are designed to capture true trait scores and study individual differences (Lord, 1958). In theory, it is possible to evaluate a person's score with reference to the available range. Consider self-esteem. Over 10 items where each receives a mark from 0 to 4, a respondent with a score of 30 may be said to have moderately high self-esteem (Rosenberg, 1965). If, however, the average score is 35, this person's score would seem relatively low. Yet, few psychologists argue that the difference of -5 between the person's score and the group mean reveals a separate trait of relative self-esteem, distinct from self-esteem proper. However, if there is variation in the group means, such as when respondents themselves estimate them, discrepancies become more likely to be seen as reflections of a separate psychological construct, as a *rei* within. Whether or not researchers reify discrepancy scores as stable internal traits or attitudes is

* Brown University, US

† Geisinger Health System, US

‡ Humboldt University, Berlin, DE

Corresponding author: Joachim I. Krueger (Joachim@brown.edu)

ultimately a theory-based decision. We counsel caution against this practice because discrepancy scores do not reveal information beyond their constituent variables. Instead, discrepancy scores are confounded with their constituents, and these confounds can be quite strong. To anticipate our conclusion, we consider discrepancy-score measures to be partially valid but impure.

In the second part of this article, we introduce an alternative, decision-theoretic, approach to the conceptualization and measurement of self-enhancement, which mitigates some of the shortcomings of conventional discrepancy scores. Though we will strongly advocate for research to explore the properties of this alternative method, we caution that at present it cannot lay claim to being a gold standard. Each extant measure testifies to the diversity of valuable perspectives on this important issue. Until a gold standard is found, researchers may wish to be mindful of the fallacy that their measure of self-enhancement is *the* measure.

Discrepancy Scores

Before addressing the psychometric challenges to discrepancy scores, let us consider some examples. A seemingly straightforward measurement of self-enhancement is obtained by estimating a person's true score T on a dimension, and then to subtract this estimate from the person's self-judgment S on that dimension. A college student might take a multiple-item test on *Art Nouveau* and then estimate the number of correct answers. The positive difference S – T (self-estimate – test score) provides a face-valid index of self-enhancement (Klar & Giladi, 1999; Moore & Kim, 2007). This method can be generalized to personality traits where T is estimated with aggregated observer judgments. To see, for example, whether the person enhances his or her self-esteem, the average peer-rated self-esteem represents an estimate of the true score. The difference S – T may represent unwarranted self-esteem, if it can be shown that the average peer knows the target person better than he or she knows him or herself (Vazire & Carlson, 2011).

This example anticipates an important distinction, namely the difference between 'error' and 'bias.' When a person's true score is estimated with the use of observer judgments, averaging them dilutes the random errors of individual observers. In the long run, these errors cancel one another out so that the average peer judgment is more accurate than the judgment of the average peer ($[T – M_P] < M[T – P_i]$) if the judgments bracket the true score T (Krueger, & Chen, 2014; Larrick, Mannes, & Soll, 2012). As the target person's self judgment S does not receive the benefit of aggregation, the person's own difference score, $S – M_P$, contains both random error and systematic bias. As the average peer-judgment also contains bias if $T – M_P \neq 0$, the discrepancy $S – M_P$ contains four elements: the target person's bias, the target person's random error, the peers' bias, and the peers' residual error. Thus, the difference score measure of self-enhancement is impure.

Now consider the study of self-enhancement in social psychology. Here, the research approach takes a different direction. Instead of asking whether a self-judgment S is higher or lower than an external reality (or social consensus) criterion T, researchers cast self-enhancement as an outcome of social comparison (Alicke, 1985; Brown, 1986). They ask whether S is higher or lower than the respondent's judgment of others. Typically, this O measure refers to a judgment of the average person, but it might be a judgment of a random or a familiar individual (Krueger, 1998). Social psychologists are mainly interested in the group effect. A focus on the group effect bypasses the need to separate those who rightly claim to be better than average from those who do so erroneously. If $M_S > M_O$, someone must have made a mistake, although it remains unknown who did. Other complications arise. The variable O does not receive the benefit of aggregation, and it is – like S – susceptible to perceptual and motivated biases. When S > O, there may be self-enhancement or other-diminishment.

A second class of discrepancy scores uses regression residuals. This approach is more common in personality-oriented research than in social psychology. Again, it is assumed that test scores or aggregated observer judgment are valid proxies of true scores T. When self-judgments S are regressed on these proxies of T, the residuals serve as measures of self-enhancement. The appeal of this measure is that it is independent of T, the drawback is a strong dependence on the simple self-judgments S. We shall argue that this dependence should give pause to those who seek to reify residual scores as a measure of an isolable trait of self-enhancement. One might ask, before having seen the results of analysis and simulation, how high the correlation between discrepancy-score measures of self-enhancement and the source variable S would have to be to conclude that the two are not differentiable. Standards of collinearity for the exclusion of predictor variables might serve as a useful heuristic here (e.g., O'Brien, 2007). Without such cautionary measures, it may seem that a separate trait has been identified particularly when the discrepancy scores show correlations with third variables of interest (e.g., happiness). Such impressions may rather be the *reification* of a mathematical term as a psychological construct (Krueger, Freestone, & MacInnis, 2013).

We explore discrepancies between S and T, which address performance "overestimation," as well as discrepancies between S and O, which address social "overplacement" (Moore & Healy, 2008), or the better-than-average effect. Note that the term overestimation invokes the idea of a judgmental *error* (i.e., misjudging one's own performance or ability), whereas the term overplacement invokes the notion of *bias* (i.e., seeing oneself as better than others). We shall argue that both, measures of overestimation and measures of overplacement, conflate judgment bias and error. We will address this issue when conceptualizing self-enhancement from a decision-theoretic perspective (Swets, Dawes, & Monahan, 2000). We will use dichotomized data to distinguish self-enhancement *bias* (the tendency to claim relative superiority) from self-enhancement *error* (a demonstrated falsity of the positively biased claim). This approach allows us to combine the study of overplacement (better than average) with overestimation (better than reality) and to acknowledge that some of those individuals who claim to be better than average are indeed better (Heck & Krueger, 2015, 2016; Krueger & Mueller, 2002; Tappin & McKay, 2016). Throughout, we refer to empirical

data collected in our lab as well as simulation studies, both of which will be available to interested researchers (available online at: https://osf.io/ky67j/?view_only=28583a8 813bd4964ada3b13ee5ee5d05).

Nonindependent Prediction

A central concern is the lack of independence of discrepancy-score measures of self-enhancement from first-order measures of the positivity of the self-concept. The two are confounded and researchers have yet to determine what degree of confounding would be considered tolerable. Regardless of the magnitude of the confound, discrepancy scores offer no incremental predictive validity beyond their constituents: they cannot predict variance in a criterion variable that is not already predicted by the discrepancy score's constituents. When asking, for example, whether self-enhancement predicts happiness, H, the correlations between S, T, and H exhaust the predictable variance. Neither the difference score S – T, nor the residual score R_S offer an increment to R^2. In other words, it cannot be said that happiness is predicted by the positivity of the self-image, S, the person's true score, T, *and* the discrepancy between S and T. The latter term is redundant, and as we argue, problematic when reified. There is, however, the possibility that the statistical interaction between S and T or polynomial transforms of each might explain additional variance (Barranti, Carlson, & Côté, 2017), but the relevance of these predictors for the theoretical construct of self-enhancement remains open to debate (Edwards, 1994; Krueger & Wright, 2011; Leising, Locke, Kurzius, & Zimmermann, 2016; Zuckerman & Knee, 1995).[1] Other concerns are that discrepancy scores lack reliability when true variability is low (Lord, 1958; Rogosa & Willett, 1983) and that they fall on arbitrary metrics that obscure "where a given score locates an individual on the underlying psychological dimension" (Blanton & Jaccard, 2006, p. 27). In the context of self-enhancement, this means that due to its relativity, a discrepancy score is difficult to evaluate without knowing the individual's standing on at least one of the constituent variables.

Despite the lack of incremental validity, many researchers have asked whether discrepancy-score measures of self-enhancement predict third variables such as happiness. This interest was inspired by Taylor and Brown's (1988) proposal that self-enhancement supports personal and social well-being. Taylor and Brown focused on research conducted with difference scores or direct measures of self-enhancement ("I am better [worse] than the average person") and self-reported measures of adjustment as outcome variables. Research with regression residuals and other outcomes, such as narcissism, yields less sanguine results (e.g., John & Robins, 1994; Paulhus, 1998). The answer to the Taylor-and-Brown hypothesis appears to depend, at least in part, on which discrepancy-score index represents self-enhancement, the valence of the predicted variable, and who judges that variable (self or observers; Krueger & Wright, 2011). We now explore some of the quantitative properties of both types of discrepancy-score measure.

The difference S – T is positively correlated with S and negatively correlated with T. McNemar (1969, p. 177)

showed how these correlations can be recovered from the correlation between the two source variables and their variances. With our notation, $r_{S-T,S} = \dfrac{s_S - r_{S,T}s_T}{\sqrt{s_S^2 + s_T^2 - 2r_{S,T}s}}$.[2]

Inspection of the numerator shows that this correlation is positive unless $s_S < r_{S,T}s_T$, that is if s_S is small. Asendorpf and Ostendorf (1998) then showed how the correlation between a difference score and a third variable can be recovered from the correlations among the three input variables and their variances. The correlation between S – T, and happiness, H, is $r_{S-T,H} = \dfrac{s_S r_{S,H} - s_T r_{T,H}}{\sqrt{s_S^2 + s_T^2 - 2s_S s_T r_{S,T}}}$.

If self-enhancement predicts happiness, it does so because simple self-judgments predict happiness better than true scores do (variances being equal). In short, difference scores contribute nothing beyond their constituents.

Now consider residual scores. When self-judgments, S, are regressed on true scores, T, the correlation between the residuals R_S and S is the ratio of the standard deviation of the residuals over the standard deviation of S, or $\dfrac{s_R}{s_S}$ (see Appendix A). This correlation cannot be negative because there are no negative variances. In contrast, bivariate regression guarantees that the correlation between the residuals and the reality variable, T, is zero (Asendorpf & Ostendorf, 1998; Leising et al., 2016). The association between the residuals and a third variable is given by the semi-partial correlation: the relationship between self-judgment and happiness controlling for true scores. In our example, self-enhancement predicts happiness after controlling for T as $r_{S,H,T} = \dfrac{r_{S,H} - r_{S,T} \cdot r_{H,T}}{\sqrt{1 - r^2_{S,T}}}$.

The first important implication is that compared with difference scores, residual scores yield stronger and more positive correlations with third variables (Krueger & Wright, 2011). For a residual-score correlation to be negative, a single correlation (here, $r_{S,H}$) needs to be smaller than the product of two correlations ($r_{S,T}$, $r_{H,T}$). In other words, the residual-score correlation is more likely to show that self-enhancement predicts happiness but also undesirable third variables such as narcissism (Paulhus, 1998). We illustrate this point with a simulation study (Simulation 1; see Appendix B) comprising 100,000 independent samples of $r_{S,T}$, $r_{T,H}$, and $r_{S,H}$, each drawn from a uniform distribution bounded by 0 and 1. In most cases (75%), we find that $r_{S,T} > r_{T,H} \times r_{S,H}$. This baseline estimate is precise as its standard error is very small. A different way to make this point is to note that the mean and the median of one correlation in the simulation are .5, whereas they are respectively .25 and .186 for the cross product of two correlations. In contrast, the difference-score correlation is not biased to show a positive self-enhancement effect on a third variable if the variances of S and T are the same.

These analytical considerations are important for an understanding of theory and research on self-enhancement. Advocates of the residual-score method emphasize the independence of R_S from the criterion variable T, while

neglecting its dependence on S. John and Robins (1994) wrote that "the residual scores represent the degree and direction of the bias that remains in the self-rankings after the behavioral-reality component has been partialed out" (p. 213). This view has since become an accepted convention. Chung, Schriber, and Robins (2016, p. 1387) write that "typically, self-reports of actual ability, standing, personality, and so on are regressed onto criterion measures of these same constructs to create an index of self-enhancement that is independent of "reality"." This convention holds that regression of self-perception, S, on reality, T, removes the valid content from self-perception, that the remaining variation can be treated as a measure of bias, and that this bias can be treated as a distinctive personality trait. Residualizing self-judgments *reifies* the concept of positive (and negative) bias as a trait (Krueger et al., 2013). Reflecting this line of thought, Leising et al. (2016) propose that the residuals' "main advantage" is that they will "by definition, be independent of the [predictors, and that they] may thus be independently interpreted" (p. 593). In other words, there is an inferential leap from noting the independence of residuals from one predictor to assuming its independent nature from another. A reified construct is one that is treated as a "thing" *sui generis*, rather than an epiphenomenon or statistical artifact. In the case of residualization, what is left after partialing a criterion cannot be considered a quantitative derivative independent of the properties inherited from its parent constructs.[3]

To explore the dependencies of the discrepancy scores empirically, we analyzed the data of 12 studies conducted with IRB approval in our laboratory between 2011 and 2016 with a total sample of 1,779 respondents (available online at https://osf.io/ky67j/?view_only=28583a8813 bd4964ada3b13ee5ee5d05). In each study, participants completed a test adapted from a trivia bank (Moore & Small, 2007). These tests were scored to obtain performance estimates T. Each respondent also estimated his or her own score, thereby providing an estimate of S, as well as the probable score of the average test taker, O. Trivia topics varied, as did the number of items per task (10, 20, or 30), and the mode of collection (Mechanical Turk; Brown University students). Two of these datasets have been published (Heck & Krueger, 2015), five served as pilot/norming studies for various trivia tasks, one was analyzed in a thesis project, three were analyzed in PRH's dissertation, and one was placed in a file drawer. We then conducted additional computer simulations to model the empirical patterns (Simulation 2; see Appendix B).

Our first concern was the dependence of discrepancy scores on simple self-judgments and the hypothesis that residual measures of self-enhancement are more strongly confounded with simple self-judgments than are difference-score measures. Over the 12 data sets, difference scores were positively but moderately correlated with self-judgments, $r_{S,S-T} = .32$ (see **Table 1** for a full correlation matrix), and accuracy was generally high (mean $r_{S,T}$ after Fisher r-Z-r transformation = .517, sem = .050). In a simulation experiment, we sampled values of S and T from normal distributions with $M = 15$ and $SD = 3$ (min = 0, max = 30) and a moderately strong accuracy correlation of $r_{S,T} = .5$. Over 100 simulated datasets, $r_{S,S-T} = .49$. Turning to residual scores, we found that, as expected, a stronger dependence on S with $r_{S,RS} = .69$ and $r_{S,RS} = .86$, respectively for the empirical and the simulated data. Relative to the typical finding in the field, our accuracy correlations were high. Zell and Krizan (2014) reported a meta-analytical estimate of $r_{S,T} = .29$. When using this modest accuracy correlation in our simulation, both discrepancy scores became noticeably more dependent on S, ($r_{S,S-T} = .605$ and $r_{S,RS} = .954$; see Leising et al. for a very similar empirical result, i.e., $r = .92$, their Table 2). Here and in similar studies, there is hardly any room left for self-enhancement to assert its psychological presence beside raw, uncorrected self-judgments.

This increase in dependency signals a general regularity such that discrepancy measures are confounded with self-judgments inasmuch as accuracy, $r_{S,T}$, is low.[4] Over our 12 empirical samples, the dependence of difference

Table 1: Complete correlation matrix over 12 samples ($N = 1,779$).

	S	O	T	S–O	S–T	$R_{S,O}$	$R_{O,S}$	$R_{S,T}$	$R_{T,S}$	$\Sigma(R_{S,T}, R_{T,S})$
O	0.72	–								
T	0.73	0.70	–							
S–O	0.46*	–0.28	0.11	–						
S–T	0.32*	–0.01	–0.42	0.45	–					
$R_{S,O}$	0.69	0.00	0.32	0.96	0.47	–				
$R_{O,S}$	0.00	–0.69	–0.25	0.89	0.34	0.72	–			
$R_{S,T}$	0.69*	0.32	0.00	0.55	0.91	0.67	0.26	–		
$R_{T,S}$	0.00	–0.25	–0.69	0.32	0.95	0.26	0.36	0.73	–	
$\Sigma(R_{S,T}, R_{T,S})$	0.36	0.03	–0.38	0.47	1.00*	0.50	0.34	0.93	0.93	–
$\Sigma(R_{S,O}, R_{O,S})$	0.39	–0.36	0.05	1.00*	0.44	0.93	0.92	0.51	0.34	0.45

Note. S = Self-judgment. O = judgment of the average other. $R_{X,Y}$ denotes a regression residual *X* computed by regressing *X* on *Y*. Asterisks denote explicit mention in the main text.

scores on self-judgments, $r_{S,S-T}$, was negatively correlated with the accuracy correlation, $r_{S,T}$, $r = -.57$. The corresponding correlation over 100 simulation experiments was $r = -.47$ (Simulation 2, see Appendix B). Residual scores also become more dependent on self-judgments as accuracy becomes lower. Across the empirical samples, the negative correlation between accuracy, $r_{S,T}$, and $r_{S,RS}$ was nearly perfect, $r = -.996$. In the simulation, relatively minor variations in $r_{S,T}$ around the central value of .5 yielded a correlation of $r_{S,RS} = -.992$. In other words, when accuracy is high, the variance of the residuals is small, which constrains their correlation with S. Conversely, when self-perceptions *lack* accuracy, the residuals are large but they closely track uncorrected self-perception. The confound between self-enhancement and the simple positivity of self-perception is mitigated when self-perceptions are quite accurate and residuals are small in size. That is – and perhaps paradoxically so – residuals can emerge as a fairly independent measure of self-enhancement only when they are small and when self-perceptions are largely error free.

To review, empirical evidence and simulation experiments show that residual-score measures of self-enhancement are more strongly confounded with the simple positivity of self-perception than are difference-score measures, and that this confound is particularly strong when self-perception lacks accuracy (i.e., when $r_{S,T}$ is low). We urge researchers to report these dependencies (not all do, e.g., Trzeniewski, Donellan, & Robins, 2008).

Inverse Regression
One reason for the appeal of residual scores is their apparent conservatism. Unlike differences, residuals are independent of one of the two source variables. If one measure (residuals) is confounded with only one variable, S, while the other measure (differences) is confounded with both S and T, the former might seem superior. As we have seen, however, this one-sided independence comes at the price of greater dependence on self-perception. Another reason for the perceived conservatism of residuals is that they are on average smaller than difference scores. In our empirical data, $M(R_S) = 2.99$ (*sem* = .058) whereas $M(S - T) = 3.34$ (*sem* = .074), a result corroborated in the simulations $M(R_S) = 2.10$, (*sem* = .026) vs. $M(S - T) = 2.39$, (*sem* = .019). If one interpretation of this effect is that residuals are purer measures of self-enhancement, another interpretation is that residuals neglect some of the self-enhancing variance in the data. Variation in T may reveal that some individuals perform worse than one would predict from their self-judgment, whereas others perform better. The former and the latter case may respectively represent a form of self-enhancement and self-effacement. We performed "inverse regressions" by predicting performance T from self-perception S and retaining the residuals. This inverse regression approach is standard practice in calibration and overconfidence research (Fischhoff, Slovic, & Lichtenstein, 1977; Gigerenzer, Hoffrage, & Kleinbölting, 1991; Larrick, Burson, & Soll, 2007). Yet, Dawes and Mulford (1996; see also Erev, Wallsten, & Budescu, 1994) recommended that regression analysis be performed both ways because they use the same data.

The choice of which variable is the predictor and which is the criterion may be arbitrary or it may be theoretically motivated, but it is not compelled by the data. The choice of regression model may both illuminate and obscure; it may even lead to opposite substantive conclusions.

When true scores, T, are regressed on self-judgments, S, the residuals R_T are independent of S, but are correlated with T with $r = \frac{s_{RT}}{s_T}$. When these residuals are reverse-scored, larger values may indicate greater self-enhancement. This method assumes that self-enhancers perform worse than their self-judgments suggest. A person with a low T score will have a somewhat higher *predicted* T score due to the law of regression. That is, this person performed more poorly than statistically expected and may therefore be considered a self-enhancer (Krueger & Mueller, 2002; Kruger & Dunning, 1999). Conversely, a person with a high T score will have a lower predicted score and may thus be considered a self-effacer. Again, self-perception accuracy affects the size of the residuals and the strength of their association with the criterion variable T. As the accuracy correlation, $r_{S,T}$, decreases, the absolute discrepancies between predicted and actual T scores increase, and the correlation between the two, $r_{T,RT}$, increases. This effect parallels the result of conventional regression. A case for a distinctive trait of self-enhancement is most easily made when self-perception is highly accurate and the residuals of T are small.

As conventional residualization captures the inequality $S > \hat{S}$ and inverse residualization captures the inequality $T < \hat{T}$, both tap into individual differences in self-enhancement, and therefore, we may expect the two types of residuals to be positively correlated. Our data show a strong association, $r = .73$,[5] and a simulation study yielded $r = .49$ (Simulation 2, see Appendix B). Without a clear theoretical argument that one type of residual score reflects self-enhancement, while the other does not, it seems prudent to use both. If both types of regression tap into the same underlying dimension, then perhaps the residuals should be combined. One option is to sum R_S and R_T (after reverse scoring) for each respondent. The necessary, though perhaps surprising, result is that the summed residuals are perfectly correlated with the simple difference scores, $S - T$ (as confirmed in our empirical data and in the simulation experiments). In other words, the claim that a particular type of residual score is distinctive (and superior) with respect to difference-scores must be supported by an argument that inverse regression is invalid or inappropriate. If this cannot be done, it would appear that difference-score measures are conceptually more appropriate than one-sided residual scores because they incorporate both the possibility that among self-enhancers S judgments may be too high and that their true scores, T, may be too low. Of course, difference scores cannot separate these two aspects of self-enhancement. At the same time, the psychometric concerns about difference scores remain. Such scores can lack reliability, obscure the magnitude of their inputs, and they never account for variance not contained in the source variables (Asendorpf & Ostendorf, 1998). Of course, researchers may opt to do both types of regression and present both residuals as well as their

sums. Yet, it seems wasteful to predict a third variable, H, from T and R_S and also predict it from S and R_T, when they might just simply predict H from S and T.

Multiple Corrections

Difference-score measures are preponderant in social-psychological studies, whereas residual-score measures find more frequent use in personality measurement studies. Social psychologists tend to subtract the respondent's own judgments or other people, O, from self-judgments, S, and personality psychologists tend to residualize self-judgments, S, by regressing them on averaged observer judgments (T in our notation). That is, there are two kinds of methodological variation over studies: the type of criterion measure (O or T) and the kind of discrepancy score (difference or residual). As we have shown, these two kinds of variation are confounded with one another.

Using difference scores, Kwan, John, Kenny, Bond, and Robins (2004) sought to combine the social-comparison perspective with the social-reality perspective by subtracting both O and T from S. This dual-difference score turns out to be highly correlated with its constituent simple difference scores and with simple self-perception, S, thereby failing to resolve the dependency problem (Heck & Krueger, 2015; Krueger & Wright, 2011). This dual-difference score does not capture the interaction between S, O, and T to predict H, but instead reflects summed differences, which obscure underlying patterns. One person may be identified as a self-enhancer because S > O and S = T, while another person is identified as a self-enhancer because S = O and S > T.

Leising et al. (2016) used multiple regression to account for social comparison and social reality. They regressed S on both T and O and retained the residualized S as the measure of self-enhancement. Having shown the limitations of regression residuals as predictors, we submit that the use of multivariate instead of bivariate residuals is a quantitative rather than qualitative adjustment. If a residual score R_S is correlated with simple S judgments, so it is correlated with S after correction for two, or more, other variables. Indeed, in their Table 2, Leising et al. (2016) report this extreme dependency ($r = .92$), raising the question once again of how much has been gained by controlling for multiple variables. Our own data show that over a total of 1,779 cases, simple self-judgments, S, were correlated with the residuals after removing variance shared with judgments of others, O, and test scores, T, at $r = .619$, which was slightly lower than the correlation between S and the residuals obtained from controlling only T ($r = .689$). The corresponding simulations yielded $r = .871$ and .820 respectively for the case of one or two controlled variables. In short, controlling for more than one variable to further purify the measurement of self-enhancement marginally reduces the dependence of self-enhancement on simple self-judgments.

All approaches considered so far are examples of a two-stage measurement and prediction paradigm. Researchers first compute discrepancy scores as differences or residuals, label them self-enhancement, and then correlate them with indices of traits of interest such as happiness. When these correlations are significant, individual differences in self-enhancement typically ignore the possibility of a third variable, presuming that the discrepancy observed plays a causal role. Such causal inferences indicate the reification of self-enhancement by arguing that, for example, people who self-enhance more will enjoy greater psychological adjustment in the future (Chung et al., 2016). There is supposed to be, in other words, a mental process captured with the residuals that contributes causally to happiness.

Humberg et al. (2017) proposed to estimate self-enhancement effects using multiple regression without following the conventional two-stage process. While retaining the simple difference score, S – T, as an individual-level measure of self-enhancement, they do not use it to predict the third variable, H. Instead, they regress H simultaneously on S and T and infer the presence of a positive self-enhancement *effect* if the regression weight of S, β1, is positive and the weight of T, β2, is negative. This multiple-regression model assumes a positive effect of self-enhancement when self-perception makes a significant positive contribution to the prediction of happiness with true performance controlled, and when at the same time performance makes a significant *negative* contribution with self-judgments controlled.

In an individual study, the regression weights β1 and β2 are independent of each other in the sense that they explain discrete portions of variance in the criterion variable. Yet, by taking both regression weights into account, the multiple-regression method resembles the approach described earlier, where both the residuals R_S and the residuals R_T are used to predict H. As we have seen, the sums of these residuals (with R_T reverse-scored) are perfectly correlated with the simple difference scores. We now ask whether using regression residuals as effect measures is different (and potentially better) than using traditional difference scores, S – T. We generated 100 simulations with 1,000 independent S, T, and H scores each (Simulation 3; see Appendix B). For each simulation, we regressed H on S and T to obtain β1 and β2 and also correlated H with the difference score S – T. As expected, the correlation between the regression weight difference, β1 – β2, and the conventional difference-score correlation $r_{S\text{-}T,H}$ was nearly perfect; $r = .999$. This correlation remained high, $r = .897$, when the source correlations reflected an empirically plausible scenario, $r_{S,T} = .5$, $r_{S,H} = .5$, $r_{T,H} = .25$. Over studies, then, the difference β1 – β2 performs much like the conventional difference score, leaving open the question of conceptual gain.

Now consider the relationship between β1 and β2 over (simulated) studies. We have seen that self-perception corrected for performance, R_S, is negatively associated with performance corrected for self-perception, R_T. As the multiple-regression approach employs the same linear model, we expect the two regression weights to be negatively correlated over samples or studies. Stronger effects of corrected self-perception should be associated with weaker or more negative effects of corrected performance. In simulations in which S, T, and H were positively intercorrelated with all $r = .5$, the correlations between β1 and β2

ranged from $-.5$ to $-.6$. This finding supports the view that regression models assess two conceptually and statistically related facets of self-enhancement: self-overestimation as reflected in R_S and $\beta1$, and underperformance as reflected in R_T and $\beta2$. In all these models, measures of overestimation are conflated with simple self-perception, and measures of underperformance are conflated with simple performance measures. Taken together, the two measures yield the same inferences as simple difference scores.

One argument central to the Humberg et al. (2017) proposal is the dependency of self-enhancement scores on simple scores reflecting the person's positive self-image, which is an issue we addressed earlier. Humberg et al. proposed to let the regression weight $\beta1$ represent the simple self-positivity effect and the difference between $\beta1$ and a negative $\beta2$ reflect the self-enhancement effect. This approach raises two concerns. First, the regression weight $\beta1$ is what most traditional researchers consider the self-enhancement effect, that is, the correlation between performance-corrected self-perception and a criterion such as H. Second, the difference between the weights $\beta1$ and $\beta2$ is dependent on $\beta1$ much like $S - T$ is dependent on S. It appears then that as much as self-enhancement is not disentangled from self-perception, so are self-enhancement effects not disentangled from simple self-perception effects. Requiring both regression weights to be significant and of opposite signs, the multiple-regression approach may provide a conservative test for the presence of a self-enhancement effect, but its critical feature, the difference between $\beta1$ and $\beta2$, is closely associated with the summed residuals R_S and R_T and the simple difference score $S - T$, and thus inherits its conceptual difficulties.

Extending pioneering work by Edwards (1994), Barranti et al. (2017) proposed Response Surface Analysis (RSA) as a general solution for research questions conventionally addressed with discrepancy scores. Applied to self-enhancement, RSA regresses a criterion variable H on two predictors S and T as well their interaction S x T and the nonlinear transforms S^2 and T^2, but does so only if the inclusion of the latter three predictors yields a statistically significant increase in the explained variance, R^2. The promise of RSA is to capture complex patterns obscured by simple discrepancies. When there is no increment in explained variance, the assessment of a self-enhancement effect on H reduces to the difference between b1 and b2, that is, the idea that S (corrected for T) predicts H better than T (corrected for S) does. In short, an RSA approach absent significant nonlinear predictors reduces to the model proposed by Humberg et al. (2017).

RSA is sophisticated and comprehensive, and given big(ger) data and strong theory, its visualization and modeling tools will likely provide new answers to questions in social and personality psychology. Still, such a multi-purpose tool may be of limited practical value in simpler environments. Lacking good theoretical grounds for the prediction of specific patterns, the analysis is rather exploratory. A five-predictor model is expensive in the number of parameter it estimates, and a complex empirical result is in greater need of cross-validation than a simpler model. Without cross-validation, a particular significant pattern runs the risk of being overfitted and may introduce false positives with incremental adjustments (Babyak, 2004; Westfall & Yarkoni, 2016). It is especially concerning that the polynomial predictors, S^2 and T^2, are highly linearly conflated with their untransformed parent variables. The numbers 1 to 10 in steps of one are correlated at $r = .97$ with their squares. Applying RSA to a dataset of our own predicting self-esteem from simple self-judgments and true scores on a trivia task yielded no significant increase in R^2 when entering the polynomial predictors; given a sophisticated tool, our own analysis could not proceed beyond S and T alone.

From Trait to Judgment: A Decision-Theoretic Alternative

We have argued that conventional measures of self-enhancement are [a] confounded with the simple positivity of self-perception, [b] tend to be reified as stable individual-differences traits, and [c] conflate a judgmental bias or response tendency with a factual error. The various types of discrepancy score reviewed here are members of a family of correlated indices. Researchers may subtract or they may residualize, and they may correct for one or more variables. None of these efforts, however, can claim to be the best, in part because of shared limitations. Decades ago, Wylie (1979) cautioned that we must "remember that, if self-reports and correlations between them are to become more clearly psychologically interpretable, *as opposed to being explainable in terms of artifactual influences*, improvements in measurement techniques and their construct validation must be effected" (p. 697, emphasis added). Today, it is still wise to remember. In the final section of this article, we sketch a decision-theoretic alternative to the conventional discrepancy-score approaches. Though we do not assert that this approach solves all the problems we have discussed, it opens another window into the conceptualization and measurement of self-enhancement.

Our point of departure is that conventional measures of self-enhancement do not separate estimates of error from estimates of bias conceptually and quantitatively (Swets et al., 2000). Many social and personality researchers treat the distinction cavalierly or ignore it altogether (Krueger & Funder, 2004). We submit that the term 'bias' be reserved for people's decision threshold, or their willingness to claim they are better than average regardless of their actual status. In contrast, the term 'error' should be reserved to the failed reality check of this claim. In other words, bias is broader than error. Error implies bias, but bias does not imply error.

To illustrate, consider the hypochondriac who is inclined (i.e., has a bias) to interpret a rash as a symptom of fatal disease, whereas a stoic has a bias to dismiss such irritations. Only when it is determined whether the dread disease is present will it be known whether the hypochondriac committed a false-positive (Type I) error or whether the stoic committed a false negative (Type II) error. It may turn out that the hypochondriac, despite being biased, made no error (true positive) and that the stoic didn't miss anything (true negative).

The distinction between the theoretically and methodologically crucial concepts of bias and error is conspicuously absent in theory and research on self-enhancement. Research routinely conflates the two. Someone who claims to be better than average, might well be. Here we observe bias in absence of error. If there is any positive association between self-judgments, S, and true scores, T, the claim to be better than most others is more likely true than false. It is essential for self-enhancement research to consider and develop this distinction. In pioneering work, Paulhus, Harms, Bruce, and Lysy (2003) introduced a measure of "overclaiming," which separates those who know much from those who claimed impossible knowledge. They assessed true and overclaimed knowledge by presenting respondents with list of items in which 20% were foils, such as "ultra-lipids" or "plates of parallax." Familiarity ratings were then analyzed with signal-detection-theory methods, which allow a separation of bias (claiming knowledge) and error (claiming impossible knowledge).

To apply the decision-theoretic framework to the interplay of self- and social judgment, we asked respondents to complete a trivia test, T, and then to predict their own performance, S, and the performance of the average other, O (Heck & Krueger, 2015). The inequality $S > O$ reflects self-enhancement bias in its social-comparison form. The difference between own performance, T, and the group median $\bar{T}$ reflects the person's actual relative standing. A self-enhancement error is indicated by the conjunction of the inequalities $S > O$ and $T < \bar{T}$ By using three types of information, S, O, and T, without combining them linearly, this method yields four types of events or decision-outcome combinations.

The decision-theoretic approach overcomes some of the limitations of conventional discrepancy scores. In addition to allowing the researcher to separate error from bias, this approach mitigates the problems of dependence and reification. In our data, self-enhancement bias (whether $S > O$) remained dependent on S, $r = .36$, but S itself does not predict self-enhancement error, $r = .02$. Simulation 2 (see Appendix) corroborated these results. S predicted bias, $r = .59$, but not error, $r = -.02$. Decision analysis emphasizes the context dependency of judgment, and it highlights the conceptual similarity of self-enhancement and overconfidence (Anderson, Brion, Moore, & Kennedy, 2012; Larrick et al., 2007). A self-enhancement error arises from the interplay of human perception and objective criteria. There is little incentive to treat an error as a 'thing' within the person who made it, that is, to reify it.

The decision-theoretic approach can be adapted for the study of individual decision-makers by using multiple tasks. Then, the probability of each of the four types of outcome can be estimated idiographically so that each person's decision space can be described by a bias and an accuracy parameter. This kind of analysis can also be performed with or without integrating the person's social comparisons (judgments of O). Such extensions will further discourage the reification of individual results or types of result.

Conclusions

We close with two reflections on the conventional discrepancy-driven measurement of self-enhancement. First, we have expressed concern about the facility with which statistical discrepancies are reified as distinctive psychological constructs. We caution against intuitive interpretations of such discrepancy scores as measures of self-enhancement. Statistical or numerical discrepancies are easily obtained, but their elevation to the status of explanatory psychological constructs requires theoretical and empirical work (*cf.* Wylie, 1979). Sometimes, the reification of a numerical discrepancy reveals a profound theoretical position, but it can degenerate into a rhetorical maneuver or a promissory note that rich theory will soon follow (Gigerenzer, 1998; Krueger et al., 2013). To illustrate, consider a case of a theoretically grounded use of difference scores. Difference scores are appealing when they arise from the subtraction of past scores from present scores. These differences can be taken to represent growth, and to be convincing, regression effects must be controlled (Fiedler & Krueger, 2012; Rulon, 1941). The notion of growth invokes causal processes whose results unfold over time. Adrian's height at age 12 can be subtracted from his height at age 16, and the difference reflects the workings of good nutrition, the growth hormone, and not having died yet. There is less room for such inferences when Adrian's height is subtracted from his father's height (Galton, 1886). Although the statistical dependencies (e.g., regression effects) are the same in both cases, the former permits a clearer causal interpretation. When judgments of others, O, or performance data, T, are subtracted from self-judgments, it might be desirable to show that judgments of the average other, O, or own performance or ability, T, cause inflations in self-perception, S, but it has been difficult to marshal evidence for this idea (Alicke, Zell, & Guenther, 2013). In self-enhancement research, the two variables are assessed concurrently, so the growth argument does not apply. Yet, many of the measure's advocates assume a motivational force that creates the differences between S and O and the residuals in S not captured by T (Alicke & Sedikides, 2011; Brown, 2012).[6]

Now consider residual scores. Residual scores $(S - \hat{S})$ *are* difference scores. They are more complex than simple differences because the subtracted variable, $\hat{S}$, is not an observed individual score but a score that is predicted from the correlation between the two source variables and from the respondent's score on the first variable, S, of this very difference. We have counseled against the reification of residual scores, and we see no such reification in most uses of multiple regression. To illustrate, consider the finding that both objective and subjective financial status predict subjective happiness. Johnson and R. F. Krueger (2006) found that subjective financial status predicted happiness even when objective status was controlled. These authors did not suggest that there is a separate and false kind of subjective financial status that predicts happiness. Instead, they concluded that subjective status, as a whole, is a particularly robust predictor. This is, we believe, a less problematic use of the regression

tool. Once covariates are controlled, the contribution of the predictor variable of interest has been cleansed as it were. This predictor now shines in a brighter light, and it is not assumed that one is now dealing with a new, conceptually unique variable representing contamination. In the ordinary usage of regression, incremental predictability is considered an asset, whereas it is considered a liability in the study of self-perception. Still, researchers should take care to guard against inflating Type I error rates when pursuing minor increments in validity (Westfall & Yarkoni, 2016).

Second, we have noted (and lamented) the proliferation of discrepancy scores dedicated to the assessment of the same psychological phenomenon. Some researchers subtract; others residualize. Some correct for one covariate; others correct for two. The shared dependency of the resulting indices on the simple self-perception, S, unites these measures in family resemblance. Some researchers are tempted to claim that their most recent version of corrected S scores is the truest of all. We find these claims unconvincing. Even more disturbing is the practice of taking one or the other measure without further discussion of its features or limitations. This approach finds modest justification in the family-resemblance argument, but it contributes to a cavalier attitude toward measurement.

Self-enhancement remains an intriguing and an important social and psychological concept. Many people, like Oscar Wilde's *Remarkable Rocket*, cultivate the notion of their own superiority. Long before Wilde, Adam Smith (1759), the rational Scotsman, agreed with this assessment and was horrified by what he saw. He observed that "we are but one of the multitude, in no respect better than any other in it; and that when we prefer ourselves so shamefully and so blindly to others, we become the proper objects of resentment, abhorrence, and execration" (chapter III, On the influence and authority of conscience). Wilde sweepingly referred to motivated self-enhancement bias, and Smith added the verdict of error. Empirical research must distinguish one from the other, for different individuals and different times.

Additional Files
The additional files for this article can be found as follows:

- **Appendix A.** Variables & Assumptions. DOI: https://doi.org/10.1525/collabra.91.s1
- **Appendix B.** Simulation Guidelines and Descriptions. DOI: https://doi.org/10.1525/collabra.91.s2

Notes
[1] A significant interaction between the predictors might reveal that the association between self-perceived scores and happiness differs depending on true scores. As a result, residual-score correlations might be computed for different subsets of respondents, a strategy that does not solve the problems of such correlations (see below).

[2] The correlation between $\hat{S}$ predicted by bivariate regression from $r_{S,T}$ and T is –1 and the correlation between S and its predicted value is +1.

[3] The status of reification, the willingness of "taking words for things" (Locke, 1849/1690, p. 104) as a fallacy remains controversial. We agree with Whitehead (1997/1925) who counseled that reification is to be avoided inasmuch as it entails a "fallacy of misplaced concreteness."

[4] McNemar's formula (see p. 3 of this article) captures this association.

[5] The correlation between the two sets of residuals is identical to the correlation between the two raw variables, $r_{S,T}$.

[6] There are additional difficulties with this claim. One difficulty is that if a constant is added to O or T to obtain an inflated S, causality cannot be shown for lack of variance.

Author Note

Competing Interests
The authors have no competing interests to declare.

References
Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology, 49*, 1621–1630. DOI: https://doi.org/10.1037/0022-3514.49.6.1621

Alicke, M. D., & Sedikides, C. (2011). *Handbook of self-enhancement and self-protection.* New York, NY: Guilford Press.

Alicke, M. D., Zell, E., & Guenther, C. L. (2013). Social self-analysis: Constructing, protecting, and enhancing the self. *Advances in Experimental Social Psychology, 48*, 173–234. DOI: https://doi.org/10.1016/B978-0-12-407188-9.00004-1

Anderson, C., Brion, S., Moore, D. A., & Kennedy, J. A. (2012). A status-enhancement account of overconfidence. *Journal of Personality and Social Psychology, 103*, 718–735. DOI: https://doi.org/10.1037/a0029395

Asendorpf, J. B., & Ostendorf, F. (1998). Is self-enhancement healthy? Conceptual, psychometric, and empirical analysis. *Journal of Personality and Social Psychology, 74*, 955–966. DOI: https://doi.org/10.1037/0022-3514.74.4.955

Babyak, M. A. (2004). What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine, 66*, 411–421. DOI: https://doi.org/10.1097/01.psy.0000127692.23278.a9

Barranti, M., Carlson, E. N., & Côté, S. (2017). How to test questions about similarity in personality and social psychological research: Description and empirical demonstration of Response Surface Analysis. *Social Psychological and Personality Science, 8*, 465–475. DOI: https://doi.org/10.1177/1948550617698204

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist, 61*, 27–41. DOI: https://doi.org/10.1037/0003-066X.61.1.27

Brown, J. D. (1986). Evaluations of self and others: Self-enhancement biases in social judgments. *Social Cognition, 4*, 353–376. DOI: https://doi.org/10.1521/soco.1986.4.4.353

Brown, J. D. (2012). Understanding the better than average effect: Motives (still) matter. *Personality and Social Psychology Bulletin, 38*, 209–219. DOI: https://doi.org/10.1177/0146167211432763

Chung, J., Schriber, R. A., & Robins, R. W. (2016). Positive illusions in the academic context: a longitudinal study of academic self-enhancement in college. *Personality and Social Psychology Bulletin, 42*, 1384–1401. DOI: https://doi.org/10.1177/0146167216662866

Dawes, R. M., & Mulford, M. (1996). The false consensus effect and overconfidence: Flaws in judgment or flaws in how we study judgment? *Organizational Behavior and Human Decision Processes, 65*, 201–211. DOI: https://doi.org/10.1006/obhd.1996.0020

Edwards, J. R. (1994). Regression analysis as an alternative to difference scores. *Journal of Management, 20*, 683–689. DOI: https://doi.org/10.1177/014920639402000311

Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous over-and underconfidence: The role of error in judgment processes. *Psychological Review, 101*, 519–527. DOI: https://doi.org/10.1037/0033-295X.101.3.519

Fiedler, K., & Krueger, J. I. (2012). More than an artifact: Regression as a theoretical construct. In Krueger, J. I. (Ed.), *Social judgment and decision-making.* New York, NY: Psychology Press, pp. 171–189.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 552–564. DOI: https://doi.org/10.1037/0096-1523.3.4.552

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland, 15*, 246–263. DOI: https://doi.org/10.2307/2841583

Gigerenzer, G. (1998). Ecological intelligence: An adaptation for frequencies. In Cummins, D.D. & Allen, C. (Eds.), *The evolution of mind.* New York, NY: Oxford University Press, pp. 9–29.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: a Brunswikian theory of confidence. *Psychological Review, 98*, 506–528. DOI: https://doi.org/10.1037/0033-295X.98.4.506

Heck, P. R., & Krueger, J. I. (2015). Self-enhancement diminished. *Journal of Experimental Psychology: General, 144,* 1003–1020. DOI: https://doi.org/10.1037/xge0000105

Heck, P. R., & Krueger, J. I. (2016). Social perception of self-enhancement bias and error. *Social Psychology, 47,* 327–339. DOI: https://doi.org/10.1027/1864-9335/a000287

Humberg, S., Dufner, M., Schönbrodt, F. D., Geukes, K., Hutteman, R., van Zalk, M. H., Denissen, J. J.,

Nestler, S., & Back, M. D. (2017). Enhanced versus simply positive: A new condition-based regression analysis to disentangle effects of self-enhancement from effects of positivity of self-view. *Journal of Personality and Social Psychology.* Advance online publication. DOI: https://doi.org/10.1037/pspp0000134

John, O. P., & Robins, R. W. (1994). Accuracy and bias in self-perception: Individual differences in self-enhancement and the role of narcissism. *Journal of Personality and Social Psychology, 66,* 206–219. DOI: https://doi.org/10.1037/0022-3514.66.1.206

Johnson, W., & Krueger, R. F. (2006). How money buys happiness: genetic and environmental processes linking finances and life satisfaction. *Journal of Personality and Social Psychology, 90*, 680–691. DOI: https://doi.org/10.1037/0022-3514.90.4.680

Klar, Y., & Giladi, E. E. (1999). Are most people happier than their peers, or are they just happy? *Personality and Social Psychology Bulletin, 25*, 586–595. DOI: https://doi.org/10.1177/0146167299025005004

Krueger, J. I. (1998). Enhancement bias in the description of self and others. *Personality and Social Psychology Bulletin, 24,* 505–516. DOI: https://doi.org/10.1177/0146167298245006

Krueger, J. I., & Chen, L. J. (2014). The first cut is the deepest: Effects of social projection and dialectical bootstrapping on judgmental accuracy. *Social Cognition, 32,* 315–335. DOI: https://doi.org/10.1521/soco.2014.32.4.315

Krueger, J. I., Freestone, D., & MacInnis, M. L. (2013). Comparisons in research and reasoning: Toward an integrative theory of social induction. *New Ideas in Psychology, 31,* 73–86. DOI: https://doi.org/10.1016/j.newideapsych.2012.11.002

Krueger, J. I., & Funder, D. C. (2004). Towards a balanced social psychology: Causes, consequences and cures for the problem-seeking approach to social behavior and cognition. *Behavioral and Brain Sciences, 27,* 313–327. DOI: https://doi.org/10.1017/S0140525X04000081

Krueger, J. I., & 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,* 180–188. DOI: https://doi.org/10.1037/0022-3514.82.2.180

Krueger, J. I., & Wright, J. C. (2011). Measurement of self-enhancement (and self-protection). In Alicke, M. D. & Sedikides, C. (Eds.), *Handbook of self-enhancement and self-protection.* New York, NY: Guilford, pp. 472–494.

Kruger, J. I., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77,* 1121–1134. DOI: https://doi.org/10.1037/0022-3514.77.6.1121

Kwan, V. S. Y., John, O. P., Kenny, D. A., Bond, M. H., & Robins, R. W. (2004). Reconceptualizing individual differences in self-enhancement bias: An interpersonal approach. *Psychological Review, 111*, 94–110. DOI: https://doi.org/10.1037/0033-295X.111.1.94

Larrick, R. P., Burson, K. A., & Soll, J. B. (2007). Social comparison and confidence: When thinking you're

better than average predicts overconfidence (and when it does not). *Organizational Behavior and Human Decision Processes, 102*, 76–94. DOI: https://doi.org/10.1016/j.obhdp.2006.10.002

Larrick, R. P., Mannes, A. E., & Soll, J. B. (2012). The social psychology of the wisdom of crowds. In Krueger, J. I. (Ed.), *Social judgment and decision making.* New York, NY: Taylor & Francis, pp. 227–242.

Leising, D., Locke, K. D., Kurzius, E., & Zimmermann, J. (2016). Quantifying the association of self-enhancement bias with self-ratings of personality and life satisfaction. *Assessment, 23*, 588–602. DOI: https://doi.org/10.1177/1073191115590852

Locke, J. (1849/1690). *An essay concerning human understanding* (30th ed.). London, England: William Tegg.

Lord, F. M. (1958). The utilization of unreliable difference scores. *Journal of Educational Psychology, 49*, 150–152. DOI: https://doi.org/10.1037/h0045456

McNemar, Q. (1969). *Psychological statistics* (4th ed.). New York, NY: Wiley.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115*, 102-117. DOI: https://doi.org/10.1037/0033-295X.115.2.502

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity, 41*, 673–690. DOI: https://doi.org/10.1007/s11135-006-9018-6

Paulhus, D. L. (1998). Interpersonal and intrapsychic adaptiveness of trait self-enhancement: A mixed blessing? *Journal of Personality and Social Psychology, 74*, 1197–1208. DOI: https://doi.org/10.1037/0022-3514.74.5.1197

Paulhus, D. L., Harms, P. D., Bruce, M. N., & Lysy, D. C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of Personality and Social Psychology, 84*, 890–904. DOI: https://doi.org/10.1037/0022-3514.84.4.890

Rogosa, D. R., & Willett, J. B. (1983). Demonstrating the reliability of the difference score in the measurement of change. *Journal of Educational Measurement, 20*, 335–343. DOI: https://doi.org/10.1111/j.1745-3984.1983.tb00211.x

Rosenberg, M. (1965). *Society and the adolescent self-image.* Princeton, NJ: Princeton University Press. DOI: https://doi.org/10.1515/9781400876136

Rulon, P. J. (1941). Problems of regression. *Harvard Educational Review, 11*, 213–223.

Russell, B. (1926/1994). *On education.* New York, NY: Psychology Press.

Smith, A. (1759). *The theory of moral sentiments.* Edinburgh, Scotland: Kincaid & Bell. DOI: https://doi.org/10.1093/oseo/instance.00042831

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1*, 1–26. DOI: https://doi.org/10.1111/1529-1006.001

Tappin, B. M., & McKay, R. T. (2016). The illusion of moral superiority. *Social Psychological and Personality Science*, advance online publication. DOI: https://doi.org/10.1177/1948550616673878

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103*, 193–210. DOI: https://doi.org/10.1037/0033-2909.103.2.193

Trzesniewski, K. H., Donnellan, M. B., & Robins, R. W. (2008). Is "Generation Me" really more narcissistic than previous generations? *Journal of Personality, 76*, 903–918. DOI: https://doi.org/10.1111/j.1467-6494.2008.00508.x

Vazire, S., & Carlson, E. N. (2011). Others sometimes know us better than we know ourselves. *Current Directions in Psychological Science*, 20, 104–108. DOI: https://doi.org/10.1177/0963721411402478

Westfall, J., & Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. *PloS one*, 11(3), e0152719. DOI: https://doi.org/10.1371/journal.pone.0152719

Whitehead, A. N. (1997/1925). *Science and the modern world.* New York, NY: Free Press.

Wilde, O. (1888/2013). *The remarkable rocket.* London, England: Penguin.

Zell, E., & Krizan, Z. (2014). Do people have insight into their abilities? A metasynthesis. *Perspectives on Psychological Science*, 9, 111–125. DOI: https://doi.org/10.1177/1745691613518075