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

Measurement of self-enhancement (and self-protection)

Joachim I. Krueger and Jack C. Wright

Brown University

Correspondence:

Joachim I. Krueger
Department of Psychology
Brown University, Box 1853
89 Waterman St.
Providence, RI 02912
Phone: (401) 863-2503
Fax: (401) 863-1300
E-mail: Joachim_Krueger@Brown.edu
Home page: http://research.brown.edu/research/profile.php?id=10378

Half of the people you know are below average.

-- Steven Wright

Sedikides and Gregg (2008, p. 102) define self-enhancers as "human beings who hold an excessively flattering view of themselves and of things associated with the self." This definition implies that self-enhancement and—to an extent—self-protection is marked by a difference between judgment and reality. Logically, this definition does not imply that everyone who has a highly positive self-image is also a self-enhancer. If Albert Einstein regarded himself a genius, hardly anyone would object. Self-enhancement means that there is a discrepancy between self-judgment and the self's true nature or value. As Alicke and Sedikides (2009, p. 28) frame the measurement issue, "to demonstrate a self-enhancement or self-protection bias is to refer a person's actions or attributions to an objective standard."

Typically, self-enhancement is measured on personality traits or other characteristics such as happiness or intelligence where a high score is desirable (Alicke & Govorun, 2005). Conversely, self-protection is measured on traits or characteristic where a low score is desirable, such as depression or vulnerability to adverse life-events (Helweg-Larsen & Shepperd, 2001; Weinstein, 1980). Although different motivational systems appear to underlie self-enhancement and self-protection (cf. Alicke & Sedikides, 2009), they are quantitative analogs of each other. Self-protection can be recast as a form of self-enhancement by reverse-coding the measure. The measurement issues reviewed in this chapter apply to both self-enhancement and self-protection, but they will be presented with reference to the former.

The researcher's task is to capture the discrepancy between self-judgment and reality. Self-judgments are easy to obtain by self-report. The challenge lies in the identification of suitable criterion (reality) variables and their reliable and valid measurement. When self-

enhancement is considered a psychological attribute of the person, its measurement faces many of the same problems arising in the measurement of any personality characteristic, only more so. How do we know that Eva is extraverted? Eva can report how extraverted she thinks she is, and a score can be calculated from her responses to a psychometric scale designed to measure extraversion. Likewise, observers can rate Eva with respect to extraversion, or they can fill out a scale with her as the target person.

Neither Eva's self-report nor the observer ratings have an automatic claim to be the superior measure. On the one hand, self-reports can benefit from the person's rich knowledge of her own behavior in a variety of social contexts (Fiedler, 1996; Krueger, Ham, & Linford, 1996; Malle & Pearce, 2001). This privileged knowledge can lead to self-enhancement. In his analysis of egotism, William James credits philosopher Adolf Horwicz with the idea that "Our own things are *fuller* for us than those of others because of the memories they awaken and the practical hopes and expectations they arouse" (James, 1890, Vol. 1, p. 327, emphasis in the original). Yet on the other hand, the person may unwittingly deceive herself or deliberately distort her judgment to manage the way others see her (Paulhus & Reid, 1991). Research in the self-enhancement tradition emphasizes the possibility of distortion while downplaying the person's advantage in accessing private information.

One common technique to overcome the small information base available to individual observers is to recruit several. The theory is that by aggregating multiple observer judgments scores can be constructed that closely represent what the person is really like. On this assumption, the discrepancy between a self-judgment and average observer judgment indicates in which direction and how strongly the person's self-view is distorted.

The aggregation of observer ratings holds two promises. First, as the sample of observers becomes larger, the random errors affecting each individual rating are gradually eliminated. Second, as the sample of observers becomes more diverse, any systematic bias arising from the observer's unique perspective is also eliminated. Ideally, a sample includes observers whose different experiences with the target person map her social world. One would want to include, for example, observers who have witnessed Eva interacting with her family and friends and other observers who know her in her working environment (Kraemer, Measelle, Ablow, Essex, Boyce, & Kupfer, 2003). Inasmuch as Eva's behavior varies over such contexts, the agreement among her observers will only be modest. A representative sampling of social contexts will likely *reduce* inter-observer agreement. Contrary to conventional wisdom, high agreement is not a good proxy for accuracy. Instead, high agreement can indicate non-representative sampling or shared observer biases.

Outline of Chapter

The observer-based approach is but one of several measurement paradigms currently in use. We begin my review by considering self-enhancement measures that do not require observer judgments. The conceptual platform for these intra-personal measures is social comparison theory (Festinger, 1954; Suls & Wheeler, 2000). From the perspective of this theory, self-enhancement is a person's perception or belief of being superior to others. We review measures devised to capture the discrepancy between judgments of self and judgments of others and examine their statistical properties. We then return to the observer-based paradigm, the conceptual platform of which is the social realist approach to personality assessment (Funder, 1995; Kenny, 2004). From this perspective, self-enhancement is

indicated by self-perception that is more favorable than the perception of the same person in the eyes of others.

A comparison of the two paradigms reveals significant similarities. Both approaches cast self-enhancement as a *discrepancy*, which entails the use of statistical difference or residual scores. At the same time, there is a meta-theoretical difference. According the social comparison approach, self-enhancement is beneficial to the person, whereas according to the social realist approach, it is a detriment to the person's well-being and adjustment. We show that careful attention to the question of "Who is judging whom with respect to which person characteristic?" can explain these divergent results.

In the final section of the chapter, We review attempts to go beyond traditional methods of measurement and analysis. Some of these alternative methods are found to be deficient, whereas others hold promise (see also Paulhus & Holden, 2008). With a gold standard for the assessment of self-enhancement still being elusive, researchers and readers are advised to be mindful of the assumptions and constraints pervading the work. Measuring self-enhancement is a far cry from sticking a meat thermometer into a turkey to see if dinner can be served.

The Social Comparison Approach

Direct Measurement

In social psychology, a popular way of measuring self-enhancement is to ask participants to judge themselves relative to other people or the average person (Suls, Lemons, & Stewart, 2002; Zuckerman & O'Loughlin, 2006). This method seeks to bypass psychometric complexities by placing the burden of measurement on the respondents themselves. In its crudest form, this "direct" approach (Alicke & Govorun, 2005) suggests

the inference that a person claiming to be better than average is guilty of a self-enhancement bias. Recalling Einstein, however, must give us pause. A person claiming to be better than average may indeed be better than average. Following classical test theory (Lord & Novick, 1968), each individual self-judgment can be regarded as a composite of a true score and bias (and random error). If true scores were known, the correlation between these scores and the comparative self-judgments would be an index of accuracy. If this correlation were positive and perfect (and if the means were the same), no one would be biased; if it were zero, a claim of being better than average would be equally likely true and false.

In social psychology, there is limited interest in the self-enhancement scores of individual research participants. Instead, it is noted that the average comparative self-judgment typically lies above the midpoint of the scale, thereby giving *prima facie* evidence of self-enhancement at the group level (Svenson, 1981). Indeed, if true scores are distributed symmetrically, it is not possible for most people to be better than average. This is not true, however, for skewed distributions. Measures of self-esteem and other self-report measures of desirable psychological attributes are highly skewed, with most people obtaining high scores (Baumeister, Campbell, Krueger, & Vohs, 2003). The self-esteem of most individuals is therefore above average (Moore & Small, 2008). Yet, only 50% of people can be above the median. To capitalize on the median's insensitivity to distributional skew, some researchers ask respondents to estimate the percentile of their score. They may ask respondents to estimate the percentage of people whose self-esteem score is lower than their own. This modified language, it is hoped, will protect respondents from confusing mean and median.

The basic finding is that estimated percentiles lie between 60% and 70% on average (Williams & Gilovich, 2008). There is some variation over participant populations, attributes

judged, and other contextual variables. For example, self-enhancement tends to be stronger for attributes relating to morality than for attributes relating to competence (Allison, Messick, & Goethals, 1989; Krueger & Acevedo, 2007), for attributes thought to be controllable rather than uncontrollable (Alicke, 1985; Heckhausen & Krueger, 1993), and for attributes whose meaning is somewhat ambiguous rather than well-defined (Dunning, Heath, & Suls, 2004; Dunning, Meyerowitz, & Holzberg, 1989). These moderator effects can be exploited in research on the cognitive and motivational sources of self-enhancement.

It is tempting to equate the bias of self-enhancement with a lack of accuracy. There is some truth to this idea. If 65% of respondents believe themselves to be above average, the maximum correlation (Φ) between belief and reality is .73. In contrast, if 95% believe themselves to be better than average (see Cross, 1977, for proof that this can happen), the maximum correlation is .23. However, one must bear in mind that the minimum correlations are constrained in analogous fashion (being -.73 and -.23 respectively for the 65% and the 95% self-enhancement effect). In short, a larger group-level self-enhancement effect narrows the window for possible accuracy correlations; it does not guarantee that judgments are overall less accurate. Likewise, extreme better-than average effects limit the variance of self-judgments, which makes it more difficult to discover systematic associations between individual differences in self-judgment and as other variables of interest, such as true scores on the focal variable or other personality variables.

The classification of individual respondents as self-enhancers or self-effacers requires independent criteria. As noted above, respondents' percentile estimates for how well they did on a test of ability can be compared (and correlated) with the percentiles associated with their test scores. Unless a test addresses an arcane subject, such as the metaphysics of Heraclitus, a

positive, though imperfect, accuracy correlation can be expected. If, for example, 65% of respondents claim to be above average, and if the accuracy correlation is .50, about a third of those claiming to be above average are not.

A positive correlation between estimated and actual percentiles permits the following predictions. The lowest scoring respondents are most likely to overestimate their scores, and the highest scoring respondents are most likely to underestimate theirs (Kruger & Dunning, 1999). With a criterion measure in play, a self-enhancer is not necessarily someone who believes to be better than average, but someone who overestimates his or her true percentile. Coupled with the overall finding that most estimates are greater than 50% it follows that the errors among the low scorers will be larger than the errors among the high scorers (Krueger & Mueller, 2002). Figure 1 (top) displays this pattern.

It is often forgotten that a bivariate distribution can be plotted in two different ways (Dawes & Mulford, 1996; Erev, Wallsten, & Budescu, 1994). When true percentiles are plotted against estimated percentiles, the respondents with the highest self-estimates are most likely to overestimate their true scores and the respondents with the lowest estimates are most likely to underestimate theirs. The absolute size of the errors is now the same (what differs is their frequency), and no one is tempted to speculate about the psychological sources of two "asymmetric errors." In Figure 1 (bottom), overestimation is indicated by a negative difference between the identity line and the regression line, whereas underestimation is indicated by a positive difference.

Nonetheless, and regardless of which plotting scheme is used, a difference measure of self-enhancement is large inasmuch as a self-rating [S] is high or a reality measure [R] is low. This result is implied by the definition of self-enhancement as S-R. When self-

judgments and criteria are assessed as an individual's percentile relative to a group, both S and R are themselves difference scores. Assuming knowledge of the underlying distributions, self-judged percentiles follow from the differences between absolute judgments about the self [S] and absolute judgments about the average person [P]. Likewise, criterion percentiles follow from the differences between true individual scores and the average score in the group [M]. In other words, the difference S-R can be expanded into (S-P) - (R-M), with S, P, and R varying over respondents, and M being a constant (Moore & Healy, 2008).

Indirect Measurement

The direct measurement approach assumes that respondents can faithfully perform an implicit social comparison. They are thought to generate estimates of how good they are or how well they do in some absolute sense. They are also assumed to generate estimates of how good the average person is or how well that person does. Finally, they are assumed evaluate the difference between these two estimates and translate it into a single composite judgment of how good they are or how well they do relative to others.

These implicit judgments and operations are rarely unpacked, but they can and they should. Why, after all, should people who are assumed to be prone to self-related biases be expected to compute their own biases rationally and reliably? Among others, Klar and Giladi (1999) decomposed the single comparative judgment by also asking participants to render absolute self-judgments and absolute judgments of the average person. Using this "indirect" approach (Alicke & Govorun, 2005), they confirmed that the difference between self-judgments, S, and judgments of the average person, P, predicted the comparative judgment, C (self judged to be better or worse relative to the average person). The question is, however, how the individual components of the difference score contribute to this result.

The correlation between the difference score S-P and the criterion C can be recovered from the variance of self-judgments, the variance of judgments of the average person, and correlations among the three variables (Cohen & Cohen, 1983). Specifically,

$$\frac{s_{\text{self}} \cdot r_{\text{self,criterion}} - s_{\text{person}} \cdot r_{\text{person,criterion}}}{\sqrt{s^2_{\text{self}} + s^2_{\text{person}} - 2 \cdot s_{\text{self}} \cdot s_{\text{person}} \cdot r_{\text{self,person}}}}.$$

Inspection of the formula shows that the difference-score correlation $r_{S\text{-}P,C}$ becomes more positive inasmuch as $s_S > s_O$ or as $r_{S,C} > r_{P,C}$. These effects are simple and lawful. Figure 2 (top) displays a plot of difference-score correlations against the natural logarithm of the ratio of standard deviations. Figure 2 (bottom) displays a plot of difference-score correlations against the natural logarithm of the ratio of correlations. Two differences are noteworthy. First, differences between $r_{S,C}$ and $r_{P,C}$ have a larger effect on $r_{S\text{-}P,C}$ than differences between s_S and s_O. The ratio of standard deviations would have to approximate infinity in order to match the maximum effect of the ratio of correlations. Second, the effect of differences between the primary correlations is harder to predict than the effect of differences between the standard deviation. Whereas the effect of the ratio of standard deviations is captured by a single ogival function, the effect of the ratio of correlations yields a family of exponential functions.

A final insight drawn from inspection of the formula for the difference-score correlation is that its value becomes more extreme as $r_{S,P}$ increases (Krueger, 2008). This latter correlation can be read as an index of accuracy on the assumption that people ought to expect similarities between themselves and others. Note, however, that a strict interpretation of accuracy would call on everyone to provide the same estimate for P (as the average person

has only one true score), in which case $r_{S,P}$ is not defined. Taken together, these lawful determinants of the size of the difference-score correlation demand that individual empirical results be interpreted with caution, as any particular value of $r_{S\text{-}P,C}$ may have arisen from a variety of different underlying patterns (see Asendorpf & Ostendorf, 1998; Krueger, 2008; Ullrich, 2009, for careful mathematical and empirical analyses).

In their empirical work on comparative judgment, Klar and Giladi (1999) confirmed (trivially) that the difference score S-P does not predict C when S and P are controlled. More importantly, they found that S was by far the best predictor of C. The correlation between P and C appeared to be spurious; its size could be estimated by multiplying $r_{S,C}$ with $r_{S,P}$ Figure 3 displays the result with sanitized though empirically representative numbers.

Several explanations of this finding have been offered (Alicke & Govorun, 2005; Chambers & Windschitl, 2004). Consider the following three. First, people may simply put more weight on their absolute self-judgments than on their judgments of the average person. This hypothesis assumes no differences between how much people know about themselves and others, only an egocentric tendency to anchor self-other comparisons on the self (Kruger, 1999). Second, it is possible that people retrieve more relevant information about the self, in part, because they focus their attention on the target of the comparison (the self) rather than the referent. According to this focalism hypothesis, comparative judgments are more closely linked to judgments about others when those others are compared with the self (Schkade & Kahneman, 1998). Third, there may be a genuine informational advantage for the self that cannot be overcome by a change of focus. If so, judgments of others are more regressive, that is, less extreme than self-judgments (Krueger, 2000; Moore & Small, 2007). Such a

restriction in variance can attenuate correlations with third variables. Recall that according to one interpretation of rationality, there *should* be little variance in judgments of the average person; the average person has only one particular value, whereas individual selves differ.

All three interpretations lead to the conclusion that a comparative judgment is problematic. As the difference score is not a suitable predictor, one might conclude that the residuals in the S judgments, after controlling for P judgments, are credible measures of self-enhancement. After all, these residuals predict the comparative judgments well. Note, however, that this line of research began with the question of *whether* the comparative measure is a good measure of self-enhancement. The results suggest that it is not. If the comparative measure is now taken to validate the residual S judgments, it begs the question of what self-enhancement is. It would not be logical to validate one measure (residual S) with a recently discredited one (the comparative judgment).

Likewise, it is no use seeking refuge in the residualized P judgments after controlling for S judgments. These residuals are independent of S. A presumed self-enhancer would be a person whose P judgment is lower than predicted on the basis of $r_{S,P}$. However, this measure cannot be validated either with reference to comparative judgments.

Despite the failure of the comparative judgment to explain unique individual differences that are not accounted for by its component variables, the measure remains of interest from a social-psychological point of view. Otten and van der Pligt (1996) found that respondents are more likely to self-enhance using a direct measure than using an indirect, componential measure. This finding suggests that the comparative measure incorporates positive self-other contrasts regardless of the basic positivity of a respondent's self-image. Such contrasts tend to be stronger when people judge the average other person than when

judging specific individuals (Alicke, Klotz, Breitenbecher, Yurak, & Vredenburg, 1995; Codol, 1975).

So far, we have seen two meanings of self-enhancement. According to one, a self-enhancer overestimates his or her true score (S-R > 0); according to another, a self-enhancer believes to be better than average (S-P > 0). How different are these measures? As both difference scores share the variable S, they must be "fundamentally related" (Larrick, Burson, & Soll, 2007, p. 78). For example, if the three variables are independent of one another and their variances are the same, the correlation between S-R and S-P can be expected to be .5.[1] Hence, compared with people with negative self-images, people with positive self-images are more likely to overestimate their true scores (e.g., on a test) and to believe they are better than others.

This simple picture changes when people's own estimated percentages of correct responses and their estimated percentages of correct responses obtained by the average person are plotted against actual percent correct. Now, there is a seemingly paradoxical result of a negative correlation. For easy tests (i.e., high actual percent correct), most people underestimate their own score but believe they did better than others (i.e., R > S > P); for difficult tests (i.e., low actual percent correct), most people overestimate their score but believe they did worse than others (P > S > R). Moore and Healy (2008) present an elegant model to resolve this conundrum. Their argument can be summarized as follows. First, a criterion variable is regressive with respect to a predictor variable unless $r = 1$ or -1. In the present context, the average self-judgments associated with the highest and the lowest actual scores are less extreme than those actual values. This is so because self-judgments include

[1] A simulation with all means set to 0 and standard deviations set to 1 shows that $r_{S\text{-}P,S\text{-}R} = .5+.5r_{P,R}$ if $r_{S,P} = r_{S,R} = 0$. Conversely, the correlation between the two difference scores decreases nonlinearly when $r_{S,P}$ or $r_{S,R}$ increase. Specifically, $r_{S\text{-}P,S\text{-}R} = -0.2394r^2 - 0.1534r + 0.4917$, where r represents $r_{S,P}$ or $r_{S,R}$.

random error in addition to genuine self-knowledge, and because these judgments are biased by prior self-knowledge extending beyond the judgment task at hand. Second, the average judgments of others associated with the highest and lowest self-judgments are less extreme than those self-judgments. This is so because people have less knowledge of others than of themselves; to make judgments of others, they fill this epistemic gap by projecting (imperfectly) from themselves. In short, self-judgments are regressive with respect to actual scores, and other judgments are more regressive still. The negative correlation between S-R and S-P over values of R is a necessary result. As actual scores rise, the better-than-average effect becomes stronger, and the overestimation of the self becomes weaker.

Interlude: The Taylor-and-Brown Hypothesis

Self-enhancement, like self-protection and self-effacement, is a psychological construct. As in the case of any such construct, its antecedents and consequences are of interest. As for the antecedents, psychologists have examined the contributions of many cognitive or motivational processes. Although a desire for a positive self-image seems all-too-human and legitimate, a bias of self-enhancement calls the person's morality and rationality into question, and thereby impugns the mechanisms underlying the bias. Self-enhancement calls up associations with collectively censured character traits such as pride, arrogance, and selfishness. Self-protection seems less objectionable for a self-protector can claim psychological self-defense as a legitimate motive. Self-effacement, although it is as much a bias as self-enhancement, raises associations with modesty and humility, and hence less opposition.

A moral subtext is difficult to avoid in social psychology (Asch, 1952; Brannigan, 2004). Yet, overt challenges to people's morality are rare. Often, the challenges address

aspects of human performance, with the message being that people are not as logical or rational as they should be. Errors, biases, and fallacies, which are among the favorite phenomena of the field, raise the specter of incoherent thinking (Dawes, 2001; Krueger & Funder, 2004). Self-enhancers, it seems, need to be educated. As Heraclitus put it during the axial age, "To extinguish hybris is more needed than to extinguish a fire."

As for the consequences, biases are a potential threat to the person's successful adaptation to life. A reasonable assumption is that biased thinking will, most of the time, degrade the accuracy of judgments. Inasmuch as accurate judgment is essential for survival, biased judgment courts negative consequences. If one believes that people learn from the consequences of their behavior, the question is "Why is there still so much self-enhancement?"

Taylor and Brown (1988) gave a surprising answer. Referring to research performed with the social comparison approach, they claimed that self-enhancement and other "illusions appear to promote other criteria of mental health, including the ability to care about others, the ability to be happy or contented, and the ability to engage in productive and creative work" (p. 193). This argument is pragmatist in the Jamesian tradition. Social beliefs can be regarded as true if they are true in their consequences, that is, if the consequences are desirable. The Taylor-and-Brown hypothesis soon met with opposition from researchers favoring a realist approach to self and social perception. The key to their opposition was the suspicion that by using a within-person measurement approach, the social comparison paradigm stacks the deck in favor of the hypothesis. To overcome this problem, self-perception ought to be compared with an external—realistic—criterion.

The Social Realist Approach

Observer-based Measurement

The backlash against the Taylor-and-Brown hypothesis was led by personality psychologists with a tradition of studying people both from the inside (self-reports) and from the outside (observer reports). These researchers favored study designs of the type introduced in the first section of this chapter. To effectively attack Taylor and Brown, however, they had to abandon a central premise of their field, namely the idea that self-perception contains valuable and valid information about the person that only that person can access. In other words, they had to agree with Taylor and Brown that discrepancies between self-judgments and criterion judgments only reveal the self's biases and distortions. Without this concession, their alternative measures of self-enhancement could not enter the contest with the intra-personal, social-psychological indices. Hence, for the purposes of the debate, the personality psychologists retreated to a reductive, behaviorist definition of personality. Armed with Hofstee's (1994, p. 155) assertion that "in a scientific context, personality is by definition a public phenomenon," they accepted inter-observer agreement as the best—if not a perfect—index of accuracy and average observer judgments as measures of reality.[2]

Despite this difference in approach, the arsenal of statistical tools remained the same during the initial phase of research. The first result was that, on average, there was little evidence for self-enhancement. The mean differences between self-judgments and aggregated observer judgments were small and inconsistent, but the vanishing of this mean-level effect received little attention. In the same way that someone wishing to ford a river has no interest in knowing that it is three feet deep *on average*, personality researchers kept the

[2] Recall that this indexing is of dubious validity from a sampling-theoretic perspective.

focus on individual differences and what they might predict. Their question was "Who self-enhances more than others and how is this related to adjustment?"

Colvin, Block, and Funder (1995) found that, compared with self-effacers, self-enhancers were described less favorably by their peers. This result contradicted the Taylor-and-Brown hypothesis. Again, unpacking the difference-score correlation is instructive. Consider again the formula $\frac{s_{self} \cdot r_{self,criterion} - s_{observer} \cdot r_{observer,criterion}}{\sqrt{s^2_{self} + s^2_{observer} - 2 \cdot s_{self} \cdot s_{observer} \cdot r_{self,observer}}}$. The difference-score correlation is negative inasmuch as $s_S < s_O$ or $r_{S,C} < r_{O,C}$. A difference in the variance of self-judgments and observer judgments may arise when observers judge multiple target persons, whereas each target judges only the self. Being able to compare target persons, observers may accentuate individual differences, whereas single target persons do not. Recall the point of the preceding section that self-judgments tend to be somewhat immune to social comparisons unless a target of comparison is explicitly provided. Even when observers consider only a single target person, they may make less regressive (i.e., more extreme) judgments than target persons themselves. The latter may be more prone to gravitate to the middle of the scale inasmuch as they try to integrate more diverse sources of information relevant for the trait at hand. Admittedly, these tendencies are counteracted by the effect of aggregation, which is that average observer judgments are regressive (i.e., have a small standard deviation) inasmuch as individual observers lack agreement with one another.

A difference between the two correlations may arise because of shared method variance. As the paradigm is staked on the idea of social reality, the criterion measures must also be provided by observers, indeed often by the same observers also judging the traits

involved in the measure of self-enhancement. Any shared bias among these observers inflates $r_{O,C}$. Defining such shared biases away with an appeal to aggregation cannot replace an empirical demonstration that such biases do not exist.

Some personality psychologists have come to distrust difference scores because of their vulnerability to regression artifacts, which has been known since the early years of the 20th Century (Thorndike, 1924). They turned to multiple regression to overcome these problems (see Edwards, 1993, for a technical description, and Zuckerman & Knee, 1996, for an application to self-enhancement research). John and Robins (1994) suggested that self-judgments be regressed on aggregated observer judgments and the residuals be retained as measures of self-enhancement. In other words, the idea was to isolate the variance in the self-judgments that could not be accounted for by social reality. John and Robins found that these residuals predicted individual differences in narcissism, that is, a positive residual correlation with an undesirable trait was consistent with the claim of Colvin et al. (1995) that self-enhancement is counter-productive. Despite using different methods, these personality psychologists were united in their rejection of the Taylor-and-Brown hypothesis.

Much as the difference-score correlation can be unpacked, so can the residual-score correlation. This correlation is the familiar semi-partial correlation between self-judgments and criterion judgments when the correlation between self-judgments and observer judgments is statistically controlled. The formula is

$$\frac{r_{\text{self,criterion}} - r_{\text{observer,criterion}} \cdot r_{\text{self,observer}}}{\sqrt{1 - r^2_{\text{self,observer}}}}.$$

It is important to note that this correlation is strongly biased to be positive. In other words, a correlation of zero is not a credible null hypothesis. Why is this so? Note that in contrast to

the case of the difference-score correlation, the inequality $r_{S,C} < r_{O,C}$ is not sufficient to yield a negative residual-score correlation. The correlation $r_{S,C}$ must be smaller than the product of $r_{O,C}$ and $r_{S,O}$. How might this happen? Although it may be expected that $r_{O,C} > r_{S,C}$ for reasons outlined above (i.e., method variance), the correlation $r_{S,O}$ would still have to be quite large.

If the expected value of the residual-score correlation is positive even before study, support for (opposition to) the Taylor-and-Brown hypothesis depends primarily on whether the criterion variable is desirable (undesirable). An unexpected result, that is, a negative residual correlation, is more likely inasmuch as the agreement between self-judgments and observer judgments is large. Taken together, these built-in constraints mean that to the extent that self-judgments appear to be accurate (high $r_{S,O}$), self-enhancement has an elevated chance of appearing dysfunctional if the criterion variable is desirable, and of appearing functional if the criterion variable is undesirable. This asymmetry is built into the statistical logic of the measure. It is not a feature of the world being studied. Unbiased tests of the Taylor-and-Brown hypothesis ought not tolerate embedded conclusions.

This leaves the option of computing the inverse residual correlation. Could one not regress observer judgments on self-judgments and correlate the residuals with a criterion? Here, a self-enhancer is a person who is judged less favorably by observers than predicted. To make this score more easily interpretable, its sign can be inverted. Again, however, the utility of this residual-score correlation is limited by the fact that it is biased to be positive, and the fact that it induces asymmetric interpretations depending on the desirability of the criterion variable (see Bereiter, 1963, for a critical analysis of regression methods).

A Hybrid Measure

It is perhaps surprising that few researchers have tried to apply different discrepancy measures (differences and residuals) to the same data (see De Los Reyes & Kazdin, 2004, for an exception). Instead, many individual investigators raise concerns about one type of measure and then employ another. Over time, cottage industries have emerged that take one measure off the shelf, accepting its validity without further inquiry. Yet, a certain discontent can be discerned. In particular, the field's inability to reach closure on the Taylor-and-Brown hypothesis has been a source of dissatisfaction. Kwan, John, Kenny, Bond, and Robins (2004) noted the differences between the social comparison approach and the social realist approach. In an effort to break the stalemate between the two, they proposed to combine them. This was a remarkable step as it required acceptance of the idea that the combination of two fallible measures will cancel rather than compound the shortcomings of each.

Kwan et al. (2004) proposed that, while the social-comparison difference of S-P and the social-reality difference of S-O both capture valid aspects of self-enhancement, each is confounded with the other. The two difference scores are correlated over respondents because they share one variable. If P and O are independent, one may expect a correlation $r_{S\text{-}P,S\text{-}O}$ of about .5, which is roughly what Kwan et al., 2004, and Sinha & Krueger, 1998, found.

To create an unconfounded index of self-enhancement, Kwan et al. (2004) adapted the social relations model (SRM; Kenny, 1994). The SRM decomposes judgments into main effects and interactions. The empirical realization of this model calls for a round-robin design, in which each person is both a target and an observer to everyone else. Judgments are then scaled relative to the mean [M] of all judgments. The target effect is given by the

difference O-M, showing whether an individual person is judged more or less favorably by others. The perceiver effect is given by the difference P-M, showing whether a particular individual person judges others more or less favorably than others do. Setting aside some corrections Kwan et al. (2004) recommended for small-groups research, a self-judgment is predicted by the two main effects and the grand mean, namely by (O-M) + (P-M) + M, or O+P-M. The difference between the actual self-judgment and its predicted value, they say, represents an idiographic interaction effect. If S-O-P+M > (<) 0, the person is a self-enhancer (self-effacer). As M is a constant, it can be ignored in correlational tests of the Taylor-and-Brown hypothesis. Hence the measure of self-enhancement is S-O-P.

The hybrid measure has been hailed as "a conceptual and methodological breakthrough" (Anderson, Srivastava, Beer, Spataro, & Chatman, 2006, p. 1098), but does it perform as intended? Is the difference S-O-P indeed an interaction effect? There is reason for doubt as one can note that the hybrid index is strikingly similar to an index obtained by simply summing the two traditional difference score, which yields 2S-O-P.

To examine the claim that the hybrid difference score captures a person-specific interaction term, it is useful to compare the score with the more familiar interaction given by analysis of variance (ANOVA). A two-way ANOVA, in which both independent variables have two manifestations, yields four cell means. After the main effects are subtracted, an interaction presents itself as a symmetrical cross-over, meaning that each residual difference between a cell mean and its predicted value is the same in absolute terms. Whether this difference is significant is a matter of statistical testing; indeed, it is the reason why ANOVA is done. This is not so in the SRM. Here, an individual residual score has no variance, and it

is unknown to what extent it is affected by measurement error. As in other discrepancy-score methods, any discrepancy is attributed to the individual's biased self-perception.

If the construal of the hybrid residual as an individual interaction term does not hold up, it remains possible, at least in theory, that this score is serviceable as a summation of the common-rater effect, S-P, and the common-target effect, S-O. However, some difficulties of interpretation remain. A person might have a score of zero because the two main effects cancel each other out. What is one to make of Eva, who self-enhances intra-individually (S-P > 0), while self-effacing inter-individually (S-O < 0)? How does she compare with Sven who self-enhances inter-individually, while self-effacing intra-individually? A pragmatic solution might be to distinguish three types of people: Those who self-enhance according to both measures, those who self-efface according to both measures, and everyone else. As S-P and S-O tend to be positively correlated, less than half of the respondents fall within the third type.

How much is gained with the use of summed difference scores? The answer is 'Not much.' The critical problem is that P has little variance. Indeed, a round-robin design using rankings makes it so because every person uses all available numbers. As the variance of P approaches zero, the hybrid term S-O-P reduces to the simple inter-individual score S-O. This is an analytical truth, and Kwan, John, Robins, and Kuang (2008) "discovered" it in an empirical study.

To illustrate some of the dependencies among the data in a round-robin matrix, consider three hypothetical patterns. Table 1 represents an idealized situation, in which the rankings of five raters are in perfect agreement. This agreement will be taken to indicate

perfect accuracy. Hence, all target effects, S-O, are zero. The perceiver effects, S-P, vary because S varies while P does not. All hybrid scores are also zero.

The numbers displayed in Table 2 are constructed from the same data, but small perturbations have been introduced. For each person, the numbers referring to two others have been switched. The consequences are revealing. From the perspective of the social reality paradigm, these perturbations constitute error. As 10 out of 25 rankings have been changed, 22 of the individual judgments now differ, if only slightly, from the aggregated O judgments. As the perturbations are small, the average O judgments retain their rank order, but their variance has become smaller. The difference S-O now has variance. As the variance in S-P has not changed, the results for S-O-P are identical to S-O. Two respondents are now classified as self-enhancers and two as self-effacers. This may be unfortunate because *the biases attributed to these respondents stem from erroneous judgments made by other observers about other targets*. A hybrid score that capitalizes on random error cannot signal a systematic interaction effect.

To introduce variance in the person variable P it is necessary to abandon the constraints of the round-robin design. One possible source of variation in P is variation in S. Psychologically, this variation is the result of social projection. The hypothetical data displayed in Table 3 are constructed from the data in Table 1 with the modification that all judgments made by Person 2 are raised by one point, and all judgments made by Person 4 are lowered by one point. The result is a change in S-O, but all hybrid scores remain zero.

If the hybrid score were taken outside of the confines of the round-robin design, the findings would only obscure these methodological problems. In a nomination design, for example, each target person recruits his or her own panel of observers, that is, people deemed

highly familiar with the target's personality (Funder, 1995). The target could rate the average person in the population or the average of his or her own observers. This approach could allow for considerable variation in P, and the hybrid score could be different from the simple differences S-O and S-P. Nonetheless, the logic of regression dictates that there is nothing left for a difference score to predict once a criterion variable has been regressed on the individual predictors (i.e., the regression weight for S-O-P = 0 once the weights for S, O, and P are estimated).

Other Methods

More Discrepancies

As we have seen, each individual difference-score measure has found its critics who recommend use of regression residuals instead. A parallel suggestion has not been made with regard to the hybrid difference score. As it has become clear that the hybrid difference-score does not reveal a statistical interaction effect, there is reason to hope that multiple regression may succeed where the ANOVA analogy failed. In multiple regression, an interaction term is expressed as the cross-product of the main effects (Cohen & Cohen, 1983).

A multiple-regression model involving an interaction term offers analytic opportunities hitherto not exploited. For example, it would be possible to test Kruger and Dunning's (1999) original hypothesis that people with low true scores make poorer social judgments than people with high true scores. When S is regressed on O, P, and OxP, this hypothesis suggests that the residuals—both positive and negative—of S relative to P are larger for individuals with low O scores than for individuals with high O scores. This result would also support Kruger and Dunning's idea that high scorers are more prone to social projection. Whereas Kruger and Dunning took social projection to be another bias, Krueger

(1998a) showed that projection increases judgmental accuracy (see also Dawes, 1989; Hoch, 1987).

Still, caution must be used in the interpretation of regression results. First, a residual of S relative to the cross-product of O and P is not an index of self-enhancement. The residual can be positive when the value of O is high and the value of P is low, or *vice versa*. Second, the asymmetries noted above still apply. Even in a full multiple regression model, outcomes of tests of the Taylor-and-Brown hypothesis depend on whether the criterion variable is desirable (e.g., self-esteem) or undesirable (narcissism) and who judges it (self or observer).

All discrepancy measures discussed so far exploit variation in judgments by the target person or variation in judgments of the target person while using normative scores for the desirability of the trait. It is assumed, for example, that everyone agrees that "sincerity" is a desirable trait and that "selfishness" is an undesirable trait. Of interest is only how strongly a person attributes the trait to the self or to others. Demonstrably, there are also individual differences in perceived trait desirability. Individuals who see themselves as sincere or selfish rate these traits as significantly more desirable than individuals who do not claim these traits for themselves.

The finding of systematic self-serving variation in trait evaluation is consistent with Dunning's theory of trait construal (Dunning et al., 1989). An idiographic self-enhancement index, which amounts to a multiple regression performed for each respondent, exploits this variation (Krueger, 1998b). Over a set of traits, self-judgments are regressed on the person's own judgments of trait desirability as well as on the group averages of self-judgments and the group averages of the desirability judgments. A self-enhancer is a person whose pattern of

overestimation and underestimation of self, relative to the social norm of self-judgments, is positively correlated with his or her pattern of overestimation and underestimation of trait desirability, relative to the social norm of desirability judgments.

Note that this "social-normative" index of self-enhancement is not an individual residual, but an idiographic regression weight (or partial correlation). In contrast to all other discrepancy measures, this index can be tested for statistical significance for each individual respondent. Perhaps more importantly, the social-normative index avoids the criterion problem that plagues the social comparison and the social reality paradigms. This method does not presume to ascertain what a person is really like in order to ascertain whether the person's self-image is too positive or too negative. The respondent neither judges other people, nor is he or she judged by others. The social-normative measure can be understood as an intra-individual association between judgmental tendencies (to rate the self high [low] and to rate traits high [low]). Other respondents only come into play by undergoing the same exercise for themselves. Their aggregated judgments furnish the normative base rates that serve as statistical controls in each individual's assessment.

Freed from the criterion problem, researchers can focus on the task of sampling traits and respondents. To help with the former, Sinha and Krueger (1998) constructed a 23-item scale based on the Big-Five Inventory (John & Donahue, 1994). Initial tests showed that, as intended, the social-normative measure is independent of the normative positivity of the self-image (i.e., $r_{S,\text{ average desirability judgments}}$). *A priori*, a person with a positive normative self-image is as likely to self-enhance (efface) as a person with a negative normative self-image. This feature should be useful in tests of the Taylor-and-Brown hypothesis, especially when the criterion trait (e.g., narcissism) is judged by observers.

Alicke, Vredenburg, Hiatt, and Govorun (2001) also proposed a within-person measure of self-enhancement. Their "better-than-myself" measure does not even require other respondents to be engaged in the same task. Each person is both a target and his or her referent. The hypothesis that people will self-enhance under this condition is so daring that even Sir Karl Popper would have liked it. Alicke et al. found that respondents rated themselves as better than average even when the estimates of relevant behaviors were their own, obtained several weeks before the comparative judgments were made. In other words, many people hold on to the idea that they are better than average even if the average person is modeled after themselves.

Recently, Preuss and Alicke (2009) proposed that meta-perceptions can be used to tap into self-enhancement. According to this approach, a self-enhancer is someone who overestimates how favorably others see him or her. Conceptually, this difference-score measure is interesting because it combines a feature of the social comparison paradigm (a judgment about another) with a feature of the social realist paradigm (a judgment made by another), while avoiding the redundancies of the hybrid measure proposed by Kwan et al. (2004). Meta-perception is logically independent of the person's private self-image, although it is empirically dependent because most people expect others to see them as favorably as they see themselves.

Beyond Discrepancies

A final family of self-enhancement measures does not involve discrepancies at all. Paulhus developed (Paulhus, 2004) and refined (Paulhus, 2002) a "Balanced Inventory of Desirably Responding" that independently assesses a person's tendency to strategically present the self favorably ("I never read sexy books or magazines.") and the tendency to

unwittingly distort the self positively ("I never regret my decisions."). These scales are suitable for the study of individual differences, but they do not reveal where self-effacement ends and self-enhancement begins. Arguably, though, the scale assessing unconscious distortions only reveals enhancement; a low score indicates the absence of enhancement but not the presence of effacement.

Paulhus also suggested that self-enhancement can be measured within a signal-detection framework (Paulhus, Harms, Bruce, & Lysy, 2003). This approach returns to the challenge of distinguishing individuals who falsely claim positive outcomes for themselves from people who do so correctly. A self-enhancer is someone with a low threshold of claiming a positive outcome. Within a person, the bias increases both false and true positives. The ratio of the two depends on how well the person's judgments are calibrated over tasks. The method thereby allows a separate assessment of bias and accuracy. When only impossible tasks are used, however, bias and inaccuracy are the same. Using a comedic adaptation of this idea, Steven Wright enjoys perplexing his audience with the challenge "Those of you who believe in psycho-kinesis, please raise my hand."

Why would people accept many false positives in their self-assessment? Error-management theory suggests that under certain conditions, people feel there is less harm in misplaced optimism than in false modesty (Haselton et al., 2009; Krueger & Mueller, 2002). Weber (1994) argued that such conditions often prevail. Her asymmetric loss function model assumes that "the more positive the outcome, the greater the cost of forgoing the beneficial effect of overestimating its likelihood and thus turning it into a self-fulfilling prophecy" (p. 230). Notice that this formulation is akin to Taylor and Brown's (1988) pragmatist interpretation of self-enhancement. Unlike Taylor and Brown, however, Weber regards the

updating of probability estimates in light of outcome valence and extremity as "a reasonable response that takes into consideration constraints that are ignored by the [standard expected utility] model" (p. 236).

Conclusions

Alicke and Govorun (2005, p. 102) wrote that "The better-than-average effect is difficult to locate in this "zoo" [. . .] of self-enhancement mechanisms." The same is true for the methods used to measure self-enhancement. Psychological characteristics are difficult to quantify. Sometimes it makes sense to aggregate different measures in hopes that each one of them contributes a piece of the picture (Campbell & Fiske, 1959). This is not so in the case of self-enhancement. Different measures imply different assumptions about who is judging whom with regard to which attributes and under what conditions (Kurt & Paulhus, 2008). There are also fundamental differences in the underlying assumptions about the statistical construction of suitable quantitative indices. It would be short-sighted to claim that "everyone's a winner," especially when different measures lead to substantively opposite conclusions (as in the debate over the Taylor-and-Brown hypothesis). Not all measures of self-enhancement can be above average.

Why is it so difficult to settle the question of proper measurement? The issue of measurement is sandwiched between the issues of hypothesis creation and hypothesis testing (Reichenbach, 1938), and sometimes it takes on features of either or both. A researcher wishing to study the trait of extraversion must have a theory of the kinds of behavior that represent extraversion and how to sample them. Scale development is theoretically driven and it is, to some extent, open to empirical pruning and refinement. If a new scale of

extraversion correlates poorly with established scales, while correlating highly with scales of neuroticism, a hypothesis has been refuted.

The dialectic of theory and data does not work as well in the case of self-enhancement. For one thing, there are no clear behavioral referents. Even if it is defined as a trait-like individual attribute, the bias of self-enhancement is measured as a judgment, or as a relationship among judgments.[3] Moreover, there is no clear way of achieving construct validity. As different measures compete for the distinction of being the true index of self-enhancement, correlations among them can neither support claims of convergent validity, nor can they demonstrate the superiority of one measure over another. Most troubling, the performance of the measures cannot be evaluated with reference to differences in predictive validity. The fate of the hybrid measure is a case in point. Acceptance or rejection of the measure cannot depend on it being more or less correlated than other discrepancy scores (most notably S-O) or with criterion variables of social success or failure. Such a decision would have to presuppose the truth or falsity of the Taylor-and-Brown hypothesis. Studies, like the one reported by Kwan et al. (2004), that seek to validate a new measure *and* test this hypothesis necessarily yield uninterpretable results. The evaluation of the measure can only be analytical, as we attempted to show in this chapter.

Ending on a realistically hopeful note, We submit that the analytical exercise presented herein has been useful. We have tried to show that there are measures of self-enhancement that show promise. Further psychometric work and independent testing of theory-driven hypotheses may yet lead to a science that will enhance insight into ourselves and the people we study.

---

[3] See, however, Holden's (1995) use of differences in response latencies to detect self-enhancing faking in self-reports.

References

Anderson, C., Srivastava, S., Beer, J. S., Spataro, S. E., & Chatman, J. A. (2006). Knowing your place: Self-perceptions of status in face-to-face groups. *Journal of Personality and Social Psychology, 91*, 1094-1110.

Alicke, M. D. (1985). Gobal self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology, 49*, 1621-1630.

Alicke, M. D., & Govorun, O. (2005). The better-than-average effect. In M. D. Alicke, D. Dunning & J. I. Krueger (Eds.), *The self in social perception* (pp. 85-106). New York, NY: Psychology Press.

Alicke, M. D., Klotz, M. L., Breitenbecher, D. L., Yurak, T. J., & Vredenburg, D. S. (1995). Personal contact, individuation, and the better-than-average effect. *Journal of Personality and Social Psychology, 68*, 804-825.

Alicke, M. D., & Sedikides, C. (2009). Self-enhancement and self-protection: What they are and what they do. *European Review of Social Psychology, 20*, 1-48.

Alicke, M. D., Vredenburg, D. S., Hiatt, M., & Govorun, O. (2001). The "better than myself" effect. *Motivation and Emotion, 25*, 7-22.

Allison, S. T., Messick, D. M., & Goethals, G. R. (1989). On being better but not smarter than others: The Muhammad Ali effect. *Social Cognition, 7*, 275-296.

Asch, S. E. (1952). *Social psychology*. New York, NY: Prentice Hall.

Asendorpf, J. B., & Ostendorf, F. (1998). Is self-enhancement healthy? Conceptual, psychometric, and empirical analysis. *Journal of Personality and Social Psychology, 74*, 955-966.

Baumeister, R. F., Campbell, J., Krueger, J. I., & Vohs, K. (2003). Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles? *Psychological Science in the Public Interest, 4*(whole No. 1).

Bereiter, C. (1963). Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 24-36). Madison, WI: University of Wisconsin Press.

Brannigan, A. (2004). *The rise and fall of social psychology*. New York, NY: De Gruyter.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

Chambers, J. R., & Windschitl, P. P. D. (2004). Biases in comparative judgments: The role of nonmotivated factors in above-average and comparative optimism effects. *Psychological Bulletin, 130*, 813-838.

Codol, J. P. (1975). On the so-called "superior conformity of the self" behavior: Twenty experimental investigations. *European Journal of Social Psychology, 5*, 457-501.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Colvin, C. R., Block, J., & Funder, D. C. (1995). Overly positive self-evaluations and personality: Negative implications for mental health. *Journal of Personality and Social Psychology, 68,* 1152-1162.

Cross, P. (1977). Not can but will college teachers be improved? *New Directions for Higher Education, 17*, 1-15.

De Los Reyes, A., & Kazdin, A. E. (2004). Measuring informant discrepancies in clinical child research. *Psychological Assessment, 16*, 330-334.

Dawes, R. M. (1989). Statistical criterion for establishing a truly false consensus effect. *Journal of Experimental Social Psychology, 25*, 1-17.

Dawes, R. M. (2001). *Everyday irrationality: How pseudo-scientists, lunatics, and the rest of us systematically fail to think rationally*. Boulder, CO: Westview Press.

Dawes, R. M., & Mulford, M. (1996). The false consensus effect and overconfidence: Flaws in judgment or flaws in how we study judgment? *Organizational Behavior & Human Decision Processes, 65,* 201-211.

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*, 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*, 1082-1090.

Edwards, J. R. (1993). Alternatives to difference scores as dependent variables in the study of congruence in organizational research. *Organizational Behavior and Human Decision Processes, 64*, 307-324.

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.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*, 117-140.

Fiedler, K. (1996). Explaining and simulating judgment biases as an aggregation phenomenon in probabilistic, multiple-cu environments. *Psychological Review, 103*, 193-214.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach.

Psychological Review, 102, 652–70.

Heckhausen, J., & Krueger, J. (1993). Developmental expectations for the self and most other people: Age-grading in three functions of social comparisons. *Developmental Psychology, 29,* 539-548.

Helweg-Larsen, M., & Shepperd, J. A. (2001). Do moderators of the optimistic bias affect personal or target risk estimates? A review of the literature. *Personality and Social Psychology Review, 5*, 74-95.

Hoch, S. J. (1987). Perceived consensus and predictive accuracy: The pros and cons of projection. *Journal of Personality and Social Psychology, 53*, 221-234.

Hofstee, W. K. B. (1994). Who should own the definition of personality? *European Journal of Personality, 8,* 149-162.

Holden, R. B. (1995). Response latency detection of fakers on personnel tests. *Canadian Journal of Behavioural Science, 27*, 343-355.

James, W. (1890). *The principles of psychology* (Vols. 1). New York, NY: Holt.

John, O. P., & Donahue, E. M. (1994). *The Big Five Inventory: Technical report of the 44-item version.* Institute of Personality Assessment and Research, University of California, Berkeley.

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.

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York, NY: Guilford.

Kenny, D. A. (2004). PERSON: A general model of interpersonal perception. *Personality and Social Psychology Review, 8*, 265-280.

Klar, Y., & Giladi E. (1999). Are most people happier than their peers, or are they just happy? *Personality and Social Psychology Bulletin, 25*, 586-595.

Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M. J. Boyce, W. T., & Kupfer, D. J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and matching contexts and perspectives. *American Journal of Psychiatry, 160,* 1566-1576.

Krueger, J. (1998a). On the perception of social consensus. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 30, pp. 163-240). San Diego, CA: Academic Press.

Krueger, J. (1998b). Enhancement bias in the description of self and others. *Personality and Social Psychology Bulletin, 24,* 505-516.

Krueger, J. (2000). The projective perception of the social world: A building block of social comparison processes. In J. Suls & L. Wheeler (Eds.), *Handbook of social comparison: Theory and research* (pp. 323-351). New York, NY: Plenum/Kluwer.

Krueger, J. I. (2008). The robust beauty of simple associations. In J. I. Krueger (Ed.), *Rationality and social responsibility: Essays in honor of Robyn M. Dawes* (pp. 111-140). New York, NY: Psychology Press.

Krueger, J. I., & Acevedo, M. (2007). Perceptions of self and other in the prisoner's dilemma: Outcome bias and evidential reasoning. *American Journal of Psychology, 120*, 593-618.

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,* 313-376.

Krueger, J., Ham, J. J., & Linford, K. M. (1996). Perceptions of behavioral consistency: Are people aware of the actor-observer effect? *Psychological Science, 7,* 259-264.

Krueger, J., & Mueller, R. A. (2002). Unskilled, unaware, or both? The contribution of social-perceptual skills and statistical regression to self-enhancement biases. *Journal of Personality and Social Psychology, 82*, 180-188.

Kruger, J. (1999). Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology, 77*, 221-232.

Kruger, J. & 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.

Kurt, A., & Paulhus, D. L. (2008). Moderators of the adaptiveness of self-enhancement: Operationalization, motivational domain, adjustment facet, and evaluator. *Journal of Research in Personality, 42*, 839-853.

Kwan, V. S., Y., John, O. P., Robins, R. W., & Kuang, L. L. (2008). Conceptualizing and and assessing self-enhancement bias: A componential approach. *Journal of Personality and Social Psychology, 94*, 1062-1077.

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.

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.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Welsley.

Malle, B. F., & Pearce, G. E. (2004). Attention to behavioral events during interaction: Two actor-observer gaps and three attempts to close them. Journal of Personality and Social Psychology, 81, 278-294.

Moore, D. A., & Healy, P J. (2008). The trouble with overconfidence. *Psychological Review, 115*, 102-117.

Moore, D. A., & Kim, T. G. (2003). Myopic social prediction and the solo comparison effect. *Journal of Personality and Social Psychology, 85*, 1121-1135.

Moore, D. A., & Small, D. A. (2007). Error and bias in comparative judgment: On being both better and worse than we think we are. *Journal of Personality and Social Psychology, 92*, 972-989.

Moore, D. A., & Small, D. (2008). When it is rational for the majority to believe that they are better than average. In J. I. Krueger (Ed.), *Rationality and social responsibility: Essays in honor of Robyn M. Dawes* (pp. 141-174). New York, NY: Psychology Press.

Otten, W., & van der Pligt, J. (1996). Contrast effects in the measurement of comparative optimism in probability judgments. *Journal of Social and Clinical Psychology, 15*, 80-101.

Paulhus, D. L. (2004). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*, 598-609.

Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun, & D. N. Jackson (Eds.), *Role of constructs in psychological and educational*

measurement (pp. 49-69). Mahwah, NJ: Lawrence Erlbaum.

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.

Paulhus, D. L., & Holden, R. D. (2008). Measuring self-enhancement: From self-report to concrete behavior. In C.R. Agnew, D.E., Carlston, D. E., W. G. Graziano, & J.R. Kelly (Eds.), *Then a miracle occurs: Focusing on behavior in social psychological theory and research* (pp.221-240). New York, NY: Oxford University Press.

Paulhus, D. L., & Reid, D. B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology, 60*, 307-317.

Preuss, G. S., & Alicke, M. D. (2009). Everybody loves me: Self-evaluations and metaperceptions of dating popularity. *Personality and Social Psychology Bulletin, 35*, 937-950.

Reichenbach, H. (1938). *Experience and prediction*. Chicago, IL: University of Chicago Press.

Schkade, D. A., & Kahneman, D. (1998). Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological Science, 9*, 340-346.

Sedikides, C., & Gregg, A. P. (2008). Self-enhancement: Food for thought. *Perspectives on Psychological Science, 3*, 102-116.

Sinha, R. R., & Krueger, J. (1998). Idiographic self-evaluation and bias. *Journal of Research in Personality, 32,* 131-155.

Suls, J., Lemons, K., & Stewart, H. L. (2002). Self-esteem, construal, and comparisons with the self, friends, and peers. *Journal of Personality and Social Psychology, 82*, 252-261.

Suls, J., & Wheeler, L., *Handbook of social comparison: Theory and research*. New York, NY: Plenum/Kluwer.

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

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103,* 193-210.

Thorndike, E. L. (1924). The influence of chance imperfections on measures upon the relationship of initial score to gain or loss. *Journal of Experimental Psychology, 7,* 225-232.

Ullrich, J. (2009). Reconsidering the "relative" in relative ingroup prototypicality. *European Journal of Social Psychology, 39*, 299-310.

Weber, E. U. (1994). From subjective probabilities to decision weights: The effect of asymmetric loss functions on the evaluation of uncertain outcomes and events. *Psychological Bulletin, 115*, 228-242.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology, 39*, 806-820.

Williams, E. F., & Gilovich, T. (2008). Do people really believe they are above average? *Journal of Experimental Social Psychology, 44*, 1121-1128.

Zuckerman, M., & Knee, C. R. (1996). The relation between overly positive self-evaluation and adjustment: A comment on Colvin, Block, and Funder (1995). *Journal of Personality and Social Psychology, 70,* 1250-1251.

Zuckerman, M., & O'Loughlin, R. E. (2006). Self-enhancement by social comparison: A prospective analysis. *Personality and Social Psychology Bulletin, 32*, 751-760.

Author Note

The author believes that this chapter is among the finest ever written about self-enhancement and its measurement. He fully realizes that the other contributors to this handbook will feel the same way about their chapters; but they are just sorely mistaken. Anyone who wishes to contest this assertion may reach the author at joachim@brown.edu. Financial support from the Alexander-von-Humboldt Foundation—who recognizes the author's brilliance—is gratefully acknowledged. The author thanks Jack Wright for many knowledge-enhancing discussions on scaling, measurement, and the detection of statistical artifacts. Mika Macinnis and Andra Geana worked magic creating the figures. David Firestone ran a simulation in MatLab, and Tony Evans, though being well above average himself, urged the author to be mindful of the average reader.

Table 1

Round-robin design take 1: Perfect consensus

Person	1	2	3	4	5	P	S-O	S-P	S-O-P+M
1	**5**	4	3	2	1	3	0	2	0
2	5	**4**	3	2	1	3	0	1	0
3	5	4	**3**	2	1	3	0	0	0
4	5	4	3	**2**	1	3	0	-1	0
5	5	4	3	2	**1**	3	0	-2	0
O	5	4	3	2	1	M=3			

Table 2

Round-robin design take 2: Some error

Person	1	2	3	4	5	P	S-O	S-P	S-O-P+M
1	**5**	4	2	3	1	3	.2	2	.2
2	5	**4**	3	1	2	3	.4	1	.4
3	5	2	**3**	4	1	3	0	0	0
4	4	5	3	**2**	1	3	-.4	-1	.4
5	5	3	4	2	**1**	3	-.2	-2	.2
O	4.8	3.6	3	2.4	1.2	M=3			

Table 3

Round-robin design take 3: Consensus plus projection

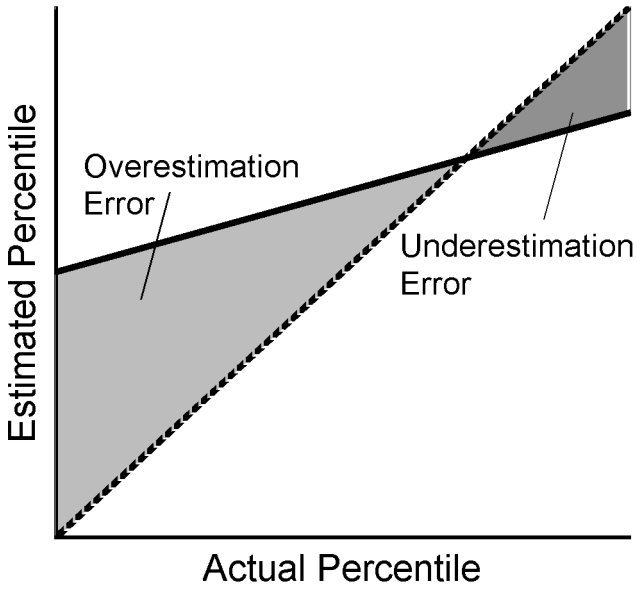
Person	1	2	3	4	5	P	S-O	S-P	S-O-P+M
1	**5**	4	3	2	1	3	0	2	0
2	6	**5**	4	3	2	4	1	1	0
3	5	4	**3**	2	1	3	0	0	0
4	4	3	2	**1**	0	2	-1	-1	0
5	5	4	3	2	**1**	3	0	-2	0
O	5	4	3	2	**1**	M=3			

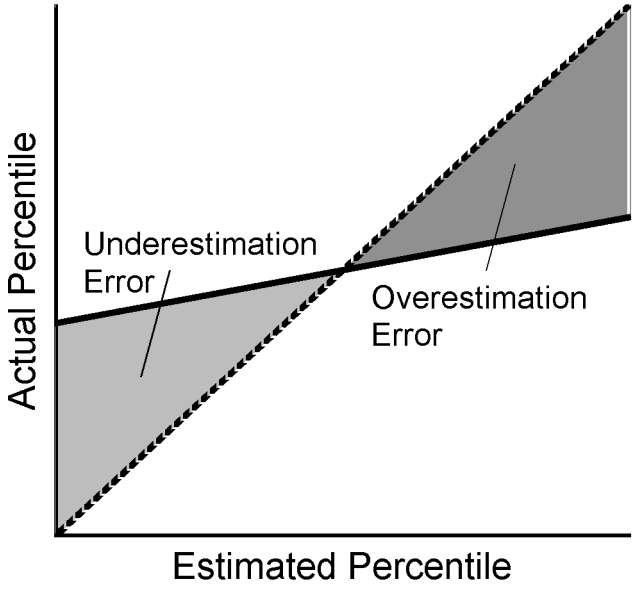
Figure Captions

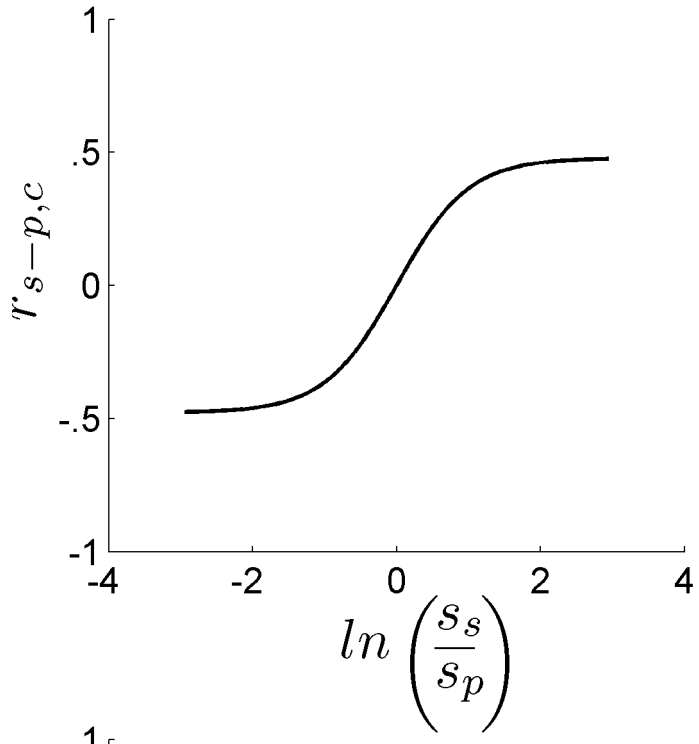
Figure 1. Predicting errors from criterion percentiles (top); predicting errors from estimated percentiles (bottom).

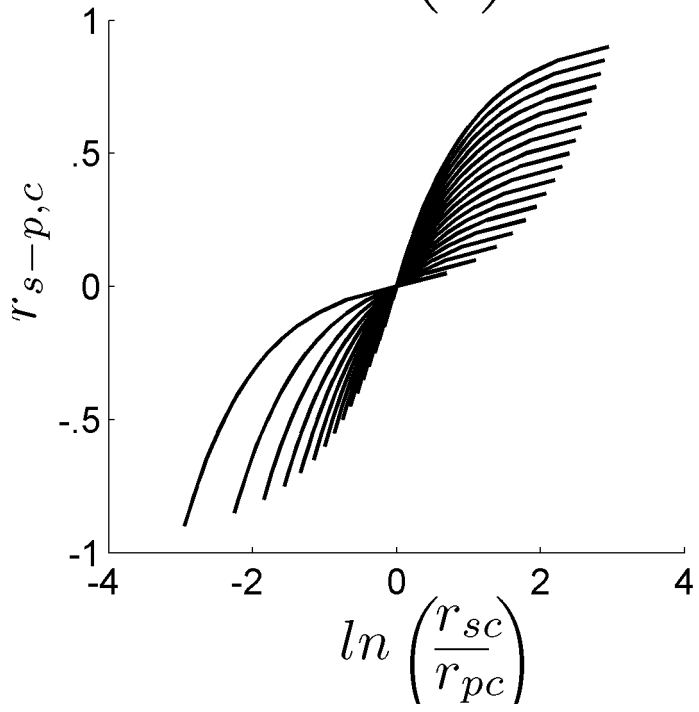
Figure 2. Predicting the difference-score correlation from the ratio of standard deviations (self-judgments over judgments of the average person) (top); predicting difference-score correlations from the ratio of correlations (self with criterion over other person with criterion) (bottom).

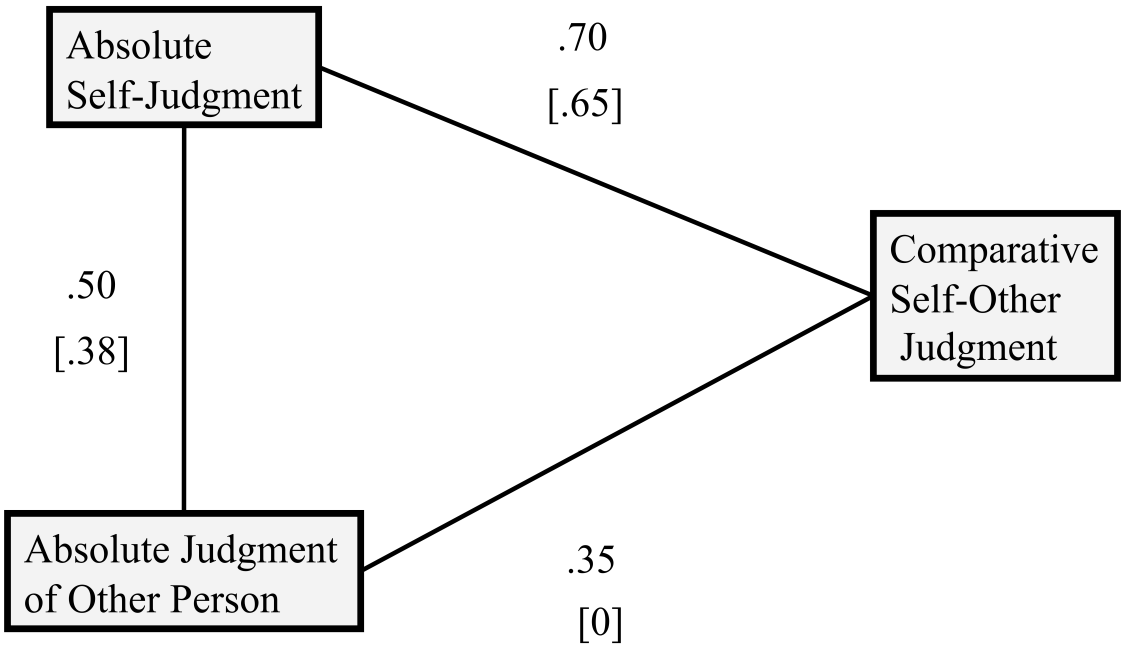
Figure 3. Comparative self-other judgments in relation to absolute self-judgments and absolute other judgments. Hypothetical (but empirically plausible) results.


Estimated Percentile
Overestimation
Error
Underestimation
Error
Actual Percentile


Actual Percentile
Underestimation
Error
Overestimation
Error
Estimated Percentile


1
.5
0
-.5
-1
-4
-2
0
2
4
$r_{s-p,c}$
$ln\left(\frac{s_s}{s_p}\right)$


1
.5
0
-.5
-1
-4
-2
0
2
4
$r_{s-p,c}$
$ln\left(\frac{r_{sc}}{r_{pc}}\right)$


Absolute
Self-Judgment
.70
[.65]
Comparative
Self-Other
Judgment
.50
[.38]
Absolute Judgment
of Other Person
.35
[0]