Is Self-Enhancement Healthy? Conceptual, Psychometric, and Empirical Analysis

Jens B. Asendorpf  
Humboldt-Universität zu Berlin

Fritz Ostendorf  
Universität Bielefeld

Whereas S. E. Taylor and J. D. Brown (1988) proposed that positive illusions foster mental health, C. R. Colvin, J. Block, and D. C. Funder (1995) presented empirical evidence that suggested the opposite: The extent to which self-ratings are more favorable than ratings by others (self-enhancement) is correlated with unfavorable personality traits (self-enhancement correlation). Our conceptual analysis shows that these propositions are not incompatible. Our psychometric analysis reveals that self-enhancement correlations are a function of the self–other agreement and the consensus of others. Our empirical analysis of self-enhancement correlations, self–other agreement, and consensus of others across 830 traits indicates that (a) self-enhancement is consistent across highly evaluative traits, but not across traits low in evaluativeness, and (b) self-enhancement correlations are minimal for non-evaluative traits and increase with increasing trait evaluativeness. We conclude that people consistently differ in their tendency to enhance their trait descriptions and that this tendency is linked to the unfavorability of their traits.

In a highly influential article, Taylor and Brown (1988) surveyed the social-psychological literature on the accuracy of social perception (including self-perception). In this survey, mental health was operationalized in most studies by the absence of low self-esteem or depression. Taylor and Brown (1988) arrived at the conclusion that mentally healthy individuals show three “pervasive, enduring, and systematic” (p. 194) illusions: unrealistically positive self-evaluations (self-enhancement), exaggerated perceptions of control or mastery, and unrealistic optimism.

This conclusion is highly provocative because it is at variance with traditional conceptions of mental health. These conceptions assert that a central characteristic of mentally healthy people is that they perceive themselves and the world around them more accurately than people plagued with mental health problems (e.g., Jahoda, 1958; Rogers, 1959).

Taylor and Brown (1988) presented the evidence that supported their conclusion in two steps for each illusion: (a) Most people (mostly university students) show the illusion, for example, self-enhancement, and (b) subjects selected for low self-esteem or mild depression, or subjects in whom a negative mood was experimentally induced, display less of the illusion, that is, show less self-enhancement.

In a trenchant critique of Taylor and Brown’s thesis, Colvin and Block (1994) suggested that the evidence for this thesis was weak at best and sometimes even contradicted Taylor and Brown’s claims. Colvin and Block’s (1994) survey of more recent studies on positive illusions and mental health also failed to lend support to Taylor and Brown’s propositions. There is no need to reiterate here the many points of critique by Colvin and Block (1994). For the conceptual analysis of the problem that we present here it is only necessary to highlight three points of critique. First, transient effects of negative mood in laboratories should not be cited as evidence for mental health effects. Second, the remaining operationalization of mental health (absence of chronic low self-esteem or depression) in the studies cited by Taylor and Brown was too specific to justify the broad term mental health. Third, many of the social–cognitive biases of normal subjects that were interpreted by Taylor and Brown (1988) as illusions are not as pervasive, enduring, and systematic as these authors assumed; instead, they are often substantially moderated and sometimes even reversed by situational and methodological factors.

For example, Alicke, Klitz, Breitenbecher, Yurak, and Vredenburg (1996) recently showed that self-enhancement, operationalized by a more favorable evaluation of self than of others, is less strong than assumed by Taylor and Brown (1988) when subjects are asked to judge close friends or relatives. Similarly, Campbell and Fehr (1990) found that self-enhancement was minimal or even reversed when self-judgments of the perfor-
performance in dyads were contrasted not with judgments of uninvolved observers but with judgments of the interaction partner. Campbell and Fehr (1990) concluded that the evidence for self-enhancement depends entirely on whose impressions are used as the criteria for accuracy.

Colvin et al. (1995) set out to refute empirically one of Taylor and Brown’s central claims: that mental health status is positively correlated with self-enhancement. They examined the relation between self-enhancement and psychological adjustment in three studies involving two different samples of young adults. Self-enhancement was operationalized by the difference in favorability between participants’ self-description and their description by others (psychologists or friends). Thus, self-enhancement was defined as stronger, the more favorable the self-description was relative to the others’ impression. The self- and other-descriptions were Q-sort profiles, and their favorability was determined by correlating each subject’s profile with a prototypic Q-sort profile that described an overly favorable personality. Subjects’ self-enhancement scores were then correlated with various judgments of personality traits or observed behaviors by another set of observers.

Regardless of the sample, sex of the subject, and the specific method of assessing self-enhancement and personality or behavior, self-enhancement correlated positively with unfavorable traits and undesirable social behaviors, and negatively with favorable traits and desirable behaviors (though with different traits or behaviors for men and women). Colvin et al. (1995) concluded from this consistent pattern of results that self-enhancement is negatively related to mental health. Thus, their conclusion apparently contradicts Taylor and Brown’s (1988) conclusion for self-enhancement. In the discussion section we will show that this contradiction is not as strong as it may appear because Taylor and Brown (1988) and Colvin et al. (1995) used very different methodologies.

There are two different interpretations of the negative correlation between self-enhancement and mental health. First, people with undesirable personality traits tend to present themselves in a more favorable light in order to avoid the disapproval of others. This form of self-presentation may be more or less deliberate and conscious (see the distinction between other-deception and self-deception by Sackeim & Gur, 1978, and Paulhus, 1994). From this view, (inter)personal problems cause self-enhancement.

Second, people who are disposed for whatever reasons to enhance their self-description behave in ways that are detrimental to their interactions with others. They evoke distancing reactions from them that are “likely to make the self-enhancers lonely and, in compensation, reinforce their narcissism” (Colvin et al., 1995, p. 1159). In the long run, they may develop unfavorable personality traits. From this view, self-enhancement causes (inter)personal problems. Both effects may operate simultaneously. Through their continuous interaction, they may feed back on each other, resulting in a vicious cycle of increasing narcissism.

Are Self-Enhancement Correlations Spurious?

Recently, Zuckerman and Knee (1996) criticized Colvin et al. (1995) on psychometric grounds. Colvin et al. (1995) correlated discrepancy scores self–other, with judgments of different others (other2). Zuckerman and Knee (1996) correctly observed that a problem may arise from the fact that other, and other2, correlate because they refer to the same persons. Zuckerman and Knee proposed to study alternatively the interaction effect Self × Other1 on Other2. As Colvin, Block, and Funder (1996) pointed out in their reply to Zuckerman and Knee (1996), this procedure is a variant of residual score analysis wherein self is regressed on other1, and the resulting residuals of self are correlated with other2. This procedure discards the baby with the bathwater because it statistically equates all individuals on other1 before the correlation between self-enhancement and other2 is examined. Such an overcorrection of meaningful differences has been observed, and criticized, for the formally similar problem of whether difference scores or residualized change scores should be used in the measurement of change (see Rogosa, Brandt, & Zimowski, 1982).

Although Zuckerman and Knee (1996) went too far with their correction procedure, correlations between self-enhancement and personality traits (self-enhancement correlations) are not easy to interpret indeed because they can be influenced by the regression to the mean in a subtle way. Consider first the more simple case that (a) the self-enhancement scores are based on the same other-assessments as the trait judgments and (b) the self- and other-judgments have equal variance. In this case, the self-enhancement correlation is negative due to the regression to the mean, and it is the more negative, the lower the self-other agreement r (self, other) is (see Equation 1):

\[ r(\text{self} - \text{other}, \text{other}) = -\sqrt{1 - r(\text{self}, \text{other})}/2. \]  

(1)

If high scores indicate a favorable trait (e.g., friendly), the regression to the mean could be erroneously interpreted as a substantive finding: Self-enhancement is negatively related to the favorability of the trait as seen by others because the more people tend to present themselves favorably in a trait, the more unfavorable their real standing is on that trait. If high scores indicate an unfavorable trait (e.g., unfriendly), the regression to the mean could be interpreted again as a meaningful psychological effect: Self-enhancement is positively related to the unfavorability of the trait as seen by others, thus negatively linked with its favorability, and this can be interpreted as discussed earlier. This observation raises the question whether Colvin et al.’s (1995) findings are amplified by such a regression to the mean, or are even completely spurious.

In defense of their substantive interpretation, Colvin et al. could reply that a regression to the mean requires that self-enhancement and trait favorability involve the same other-assessment (see Equation 1). In every one of the three studies by Colvin et al. (1995), however, self-enhancement and trait favorability were based on different other-assessments. However, the other-assessments were correlated, of course, because they referred to the same individuals. In studies 1 and 2, the favorability scores of the two independent other-assessments correlated between \( r = .58 \) and \( r = .64 \) (C. R. Colvin, personal communication, July 1996).

Because the two other-assessments were correlated, the regression to the mean for the one was transferred to some extent to the other. That this transfer exists is easy to understand without
The discrepancy between the self-view and the mean judgment of the first set of observers will be higher the lower this mean judgment is. Because the observer judgments are correlated between the two sets of observers, this will also apply to the mean judgment of the second set of observers to some extent. It is also obvious that this transfer will be stronger the higher the observer judgments are correlated. In the Appendix we prove the following equation for variables with the same variance:

\[
r(\text{self}, \text{other}_1, \text{other}_2) = \frac{r(\text{self}, \text{other}_2) - r(\text{other}_1, \text{other}_2)}{\sqrt{2 \cdot (1 - r(\text{self}, \text{other}_1))}}.
\]  

Thus, for self- and other-descriptions of equal variance, the self-enhancement correlation is a function of the two possible self-other agreements and the consensus among others. Note, however, that this psychometric truth about self-enhancement correlations does not imply that these correlations are spurious. The self-other agreements, the consensus among others, and differences between them are empirical facts that convey psychological meaning. Because the denominator in Equation 2 is always positive, the self-enhancement correlation will be negative to the extent that the self-other agreement is lower than the consensus among others. Contrary to the regression to the mean, the self-enhancement correlation can be positive if the self-other agreement in the numerator in Equation 2 is higher than the consensus among others.

In the more general case that the self- and other-assessments have unequal variances, the correlations on the right side of Equation 2 are weighted by these variances (see Appendix, Formula A11). Although the resulting formula is complex, it can be shown that the self-enhancement correlation increases to the extent that the self-other agreement becomes smaller relative to the consensus of others, and that the self-assessment becomes less variable than the other-assessments; the latter effect becomes stronger with increasing self-other agreement (see Appendix, Figure A1). The self-enhancement correlation can be as strong as \( r = -.60 \) if (a) the consensus among others is very high \( (r = .80) \), (b) the self-other agreement is much lower \( (r = .50) \), and (c) the standard deviation for the self-assessment is only half as large as the standard deviation for the other-assessments. This is an extreme case, though; in most cases, not all three conditions are met, and the self-enhancement correlation will be less strong.

To summarize, our psychometric analysis shows that self-enhancement correlations are not spurious. Instead, they depend on two psychologically meaningful properties of trait judgments: (a) the extent to which the self-other agreement is smaller than the consensus of others, and (b) the extent to which the self-judgments are less variable than the other-judgments. These two properties can vary across different traits. Therefore, the self-enhancement correlations can also vary across traits.

Self-Enhancement Correlations and Trait Evaluativeness

Because self-enhancement correlations covary across different traits with the self-other agreement, the consensus of others, and with the variability of the self- and other-judgments, the covariation of these properties of trait judgments across different traits can be studied empirically to get some idea about the magnitude of the association between traits and self-enhancement and how it is linked to the agreement, the consensus, and the variability of the trait judgments.

Some of these cross-trait relations have already been examined. John and Robins (1993) related the evaluativeness of different traits (the degree to which they are considered favorable or unfavorable in the culture under investigation) to the agreement in these traits. A sample of 76 trait adjectives were judged by self and four peers. Peer-peer consensus ranged from .02 to .50, with a mean of .25; self-peer agreement ranged from -.07 to .55, with a mean of .19. Evaluative traits showed a lower self-peer agreement than peer consensus, whereas no such difference was found for nonevaluative traits. This is exactly what one would expect from the perspective of self-enhancement: Self-enhancement decreased the self-peer agreement, but only for evaluative traits because no self-enhancement is expected for nonevaluative traits.

More importantly, it is possible to test in such a cross-trait analysis a fundamental assumption about self-enhancement directly: Self-enhancement in a trait is bound to the evaluativeness of this trait, that is, the extent to which the trait is considered to be favorable or unfavorable in the culture under investigation. For evaluatively neutral traits (traits that are considered neither favorable nor unfavorable), the notion of self-enhancement is meaningless indeed. Self-enhancement in a trait requires that self- and other-judgments (or other personality criteria) can be interpreted as favorable or unfavorable. From this fundamental assumption, four testable cross-trait hypotheses can be delineated.

**Hypothesis 1:** Self-enhancement correlations are zero for evaluatively neutral traits.

**Hypothesis 2:** Self-enhancement correlations become more negative with increasing trait evaluativeness (the extent to which a trait is judged as favorable or unfavorable).

**Hypothesis 3:** Interindividual differences in self-enhancement are consistent across highly evaluative traits.

**Hypothesis 4:** The consistency of interindividual differences in self-enhancement across different traits increases with increasing trait evaluativeness.

The first two hypotheses can be derived from the fundamental assumption and from Colvin et al.'s (1995) findings as follows. Self-enhancement correlations that arise from self-enhancement rather than from other factors, including indirect effects of the regression to the mean, require systematic interindividual differences in the tendency to enhance one's self-description in a trait. For evaluatively neutral traits, these interindividual differences are expected to be zero because there is nothing to enhance for any person. With increasing trait evaluativeness, interindividual differences in self-enhancement are expected to increase, and the self-enhancement correlations become stronger. According to Colvin et al. (1995), they should become more negative.

The two other hypotheses follow also from the fundamental assumption. If two traits are similarly favorable or unfavorable, they stimulate self-enhancement to a similar degree. Consequently, self-judgments in these two traits share one common...
underlying factor: interindividual differences in self-enhancement. The more two traits are evaluative, the more they share variance due to self-enhancement, and the more self-enhancement is consistent between the traits, regardless of the specific other meaning of the traits.

This approach is one way to disentangle substantive self-enhancement effects from potential statistical artifacts. Trait evaluativeness is a substantive concept, based on the social meaning of traits. If it can be shown that self-enhancement effects are absent for nonevaluative traits but increase in strength with increasing trait evaluativeness, this would strongly support substantive interpretations of self-enhancement effects.

Therefore we set out to shed more light on the self-enhancement issue by testing these four hypotheses in a sample of 830 trait items that were rated by self and three acquaintances; favorability ratings were also available for all traits (Ostendorf, 1990). Because each person was judged on each trait by three different acquaintances, we defined self-enhancement by contrasting the self-ratings with the ratings of one acquaintance, and then correlated these self-enhancement scores with the judgments of the two remaining acquaintances. Then we correlated the resulting self-enhancement correlations with trait evaluativeness across the 830 trait items to test our Hypotheses 1 and 2.

Furthermore, we tested our Hypotheses 3 and 4 by analyzing the internal consistency of self-enhancement for 41 scales, each consisting of 20 items of equivalent favorability. We expected that the internal consistency of these scales would be high for highly evaluative scales and would increase with increasing trait evaluativeness.

Finally, we could capitalize on the aggregation principle (Epstein, 1983) by studying the relation between scale evaluativeness and the self-enhancement correlation not only at the level of the 830 items but also at the level of the 41 scales. We expected that this relation would be stronger at the level of scales because the scales maximize the interindividual variance due to evaluativeness, and minimize the variance due to other factors, particularly measurement error.

Method

Participants

Participants were 511 individuals (194 males, 285 females, 32 did not report their sex) ranging in age from 15 to 85 years (M = 29.89 years, SD = 12.61 years). They were recruited through announcements in the local newspapers of 12 German and one Austrian university cities. Testing took place at the local universities. Each participant was accompanied by three acquaintances who provided the other-ratings. To secure these, we computed for each item the mean self-judgment and the self-ratings of the two remaining acquaintances. Then we correlated the resulting self-enhancement correlations with trait evaluativeness across the 830 trait items to test our Hypotheses 1 and 2.

To study the average self-enhancement across the 830 items, we computed their absolute value. As Figure 1, right panel, indicates, the resulting distribution was negatively skewed: There were more evaluative than nonevaluative items. Such a pattern has also been found for large samples of English and Dutch terms, and has been interpreted as support for the hypothesis that trait terms in natural languages are ideal-based and goal-derived social categories (see Borkenau, 1990).

Average Self-Enhancement

To study the average self-enhancement across the 830 items, we computed for each item the mean self-judgment and the mean other-judgment, both separately for each acquaintance and aggregated over the three acquaintances. Because the results in this section were virtually identical for the different acquaintances, we present here only the results for the judgments averaged over the three acquaintances. First, we correlated the mean self-judgment and the mean other-judgment with the item's favorability across the 830 items.

As expected (see, e.g., Edwards, 1953), the favorability of the items was strongly related to both the mean self-rating (r = .90) and the mean other-rating (r = .94). The self-ratings increased somewhat less than the other-ratings; the differences mean(self) - mean(others) correlated .57 with item favorability. The regression lines in Figure 2 indicate an average self-devaluation effect: For unfavorable items, the subjects rated...
themselves higher than their acquaintances on average, whereas they rated themselves somewhat lower than their acquaintances on favorable items.

Hypotheses 1 and 2: Item Level

For each item, there were three other-ratings and thus six different possibilities to compute a self-enhancement correlation for this item, for example, \( r(\text{self} - \text{other}_1, \text{other}_2) \), \( r(\text{self} - \text{other}_1, \text{other}_3) \), \( \ldots \). We averaged these six correlations for each item by using Fisher’s \( r \) to \( z \) transformation for correlations and thus obtained one robust estimate for the self-enhancement correlation of that item. These correlations varied between -.17 and .10, with a mean of -.02. Thus, their effect size was generally quite small. Due to the large number of items, however, the mean self-enhancement correlation deviated significantly from zero, \( t(829) = 17.55, p < .001 \).

To test Hypotheses 1 and 2, we correlated these 830 self-enhancement correlations with the evaluativeness of the items. This cross-item correlation was \(-.32 (p < .001)\), confirming Hypothesis 2. As Figure 3 indicates, the average self-enhancement correlation was very close to zero for nonevaluative items and then decreased to an average of approximately \(-.05 \) for the most evaluative items.

To test Hypothesis 1 (whether the self-enhancement correlations deviated significantly from zero for nonevaluative items), we computed the self-enhancement correlations for the 27 items with an evaluativeness score smaller than 0.1. These correlations varied between \(-0.3 \) and \( .05 \) \((M = 0.00, SD = .03, t < 1, ns)\). Thus, our first two hypotheses were both confirmed. However, the variability of the self-enhancement correlations at a similar level of item evaluativeness was substantial.

Separate analyses for favorable items (\( z \)-score for favorability above 0) and unfavorable items (\( z \)-score for favorability below 0) yielded negative correlations between the evaluativeness and the self-enhancement correlation of the items in both cases: For favorable items, \( r(412) = -.45, p < .001 \); for unfavorable items, \( r(418) = -.18, p < .001 \). The correlation for the favorable items was significantly more negative than the correlation for the unfavorable items, \( z = 4.35, p < .001 \). Thus, the self-enhancement correlation was more strongly related to cross-item variations in favorability than to cross-item variations in unfavorability.

Hypotheses 1 and 2: Scale Level

Both the size of the self-enhancement correlations and the size of the correlation between the self-enhancement correlations and the evaluativeness of the items were small. Much higher correlations cannot be expected because these analyses were
based on judgments of individual items by individual raters (see, e.g., Epstein, 1983). Therefore we tried to reduce error variance by aggregating ratings across items of similar favorability.

We randomly dropped 10 items from the pool of 830 items and then constructed 41 scales consisting of 20 items of similar favorability. The 20 most favorable items formed the first scale, the 20 most favorable items of the remaining item pool made up the second scale, and so forth. Because the only factor common to all items of a scale was their favorability, each scale aggregated the effect of a particular level of favorability on the ratings of its items and minimized the effect of all other factors that determined the ratings, including measurement error.

Using these 41 scales, we replicated all major findings for the items at the item level, but with much higher effect sizes. The distribution of the scales' favorability was very similar to the distribution of the items' favorability. Therefore, we again z-transformed the scales' favorability scores and used the absolute z-score as a measure of the scales' evaluativeness. The mean favorability of the scales was strongly related to both the mean self-ratings (r = .992) and the mean other-ratings (r = .996). The self-ratings increased somewhat less than the other-ratings; the differences mean (self) − mean (others) correlated strongly (r = −.935) with the mean favorability of the scales. The regression lines again indicated a general self-devaluation effect.

The self-enhancement correlations for the 41 scales varied between −.20 and .04, with a mean of −.10. Thus, their effect size was still small. The mean self-enhancement correlation deviated significantly from zero, t(40) = 13.25, p < .001. The correlation between the scales' evaluativeness and the self-enhancement correlation was −.72, a clear linear relation (see Figure 4). The self-enhancement correlation for the least evaluative scale (evaluativeness 0.1) was .00. Thus, again our first two hypotheses were confirmed.

**Agreement, Consensus, and the Self-Enhancement Correlation**

As we have shown in our psychometric analysis in the introductory section, the self-enhancement correlation of a trait is a function of four characteristics of the trait: self—other agreement, consensus among others, standard deviation of the self-ratings, and standard deviation of the other-ratings. Table 1 presents the mean of these trait characteristics, their variability, and their correlations with trait evaluativeness for the 830 items and the 41 scales.

Table 1 indicates that the average agreement between self and one judge was somewhat lower than the average consensus of one judge with another judge; the difference was highly significant both for items, t(829) = 14.81, p < .001, and for scales, t(40) = 11.86, p < .001. This difference confirmed the finding by John and Robins (1993), but with a much larger number of traits.

The self—other agreement was less positively correlated with evaluativeness than the consensus among others, both for items and for scales; consequently, the difference between agreement and consensus was significantly negatively correlated with evaluativeness for both items, r(830) = −.35, p < .001, and scales, r(41) = −.68, p < .001. This effect is consistent with the self-enhancement view: The more evaluative a trait is, the more self—other agreement is diminished by interindividual differences in self-enhancement. This increasing effect, in turn, decreases self—other agreement relative to the consensus among others. Independent of this effect, the consensus among others increased with increasing trait evaluativeness (for items, r = .12, p < .001; for scales, r = .77, p < .001). This between-trait effect may be due, for example, to a higher monitoring of evaluative traits by others because such traits are more important to them.

**Hypotheses 3 and 4**

We tested Hypotheses 3 and 4 on the consistency of self-enhancement across different traits as follows. First, we computed for each participant and item the difference between the self-rating and the average other-rating. For favorable scales, this difference score is a measure of self-enhancement; for unfavorable scales, the reversed difference is a measure of self-enhancement. Second, we computed the internal consistency of these individual self-enhancement scores for each 20-item scale (Cronbach’s α).
Figure 5. Correlation between scale evaluativeness and the reliability of 41 self-enhancement scales.

Figure 5 shows that the internal consistencies of these self-enhancement scales were closely related to the evaluativeness of the scales ($r = .95$) and were satisfactory for the more evaluative scales ($\alpha > .80$ for the scales with evaluativeness scores in the upper third of the distribution). Thus, Hypotheses 3 and 4 were also both confirmed. In particular, it was possible to measure self-enhancement with the more evaluative scales with sufficient reliability. This last analysis can be conceived as a rational scale construction for the measurement of self-enhancement that capitalizes solely on the knowledge of the favorability of items and the aggregation principle.

Discussion

Is self-enhancement healthy? Colvin et al. (1995) found negative correlations between individuals’ overall self-enhancement of their personality and the overall favorability of their personality as judged by others. They interpreted these negative self-enhancement correlations as substantive, motivational effects: Those with an unfavorable standing on the trait may be more highly motivated to present themselves more favorably, or tendencies for individuals to enhance their self-descriptions correlate with unfavorable trait attributions by others because self-enhancement may evoke distancing reactions from them.

Psychometric Analysis

Zuckerman and Knee (1996) proposed that Colvin et al.’s (1995) correlational approach may be problematic on psychometric grounds because they correlated difference scores between two variables with a third variable that is correlated with the components of the difference score. As Colvin et al. (1996) pointed out in their reply to this critique and to the solution proposed by Zuckerman and Knee (1996), Zuckerman and Knee’s (1996) proposal leads to an overcorrection of meaningful differences.

Our psychometric analysis went beyond both Zuckerman and Knee (1996) and Colvin et al. (1996). We showed that the self-enhancement correlation is a function of the self—other agreement, the consensus of others, and the variance of the self- and other-judgments. These are properties of trait judgments that have substantive meaning and can vary across different traits. Thus, self-enhancement correlations are not spurious. They reflect meaningful psychological phenomena that also affect the self—other agreements, the consensus among others, and the variances of the assessments. However, our psychometric analysis also made it clear that self-enhancement correlations capture the combination of quite diverse influences on self- and other-assessments. For example, any factor that changes the difference between self—other agreement and consensus among others affects the size of the self-enhancement correlation.

Factors that increase the self-enhancement correlation can include the following: Others may know people only from the same restricted range of situations or share specific halo effects (e.g., related to physical attractiveness) or prejudices (increasing their consensus); others know different people from different kinds of situations (increasing the variance of the other-assessments). Factors that decrease the self-enhancement correlation include: Other-assessments are heterogeneous because judgments from different kinds of judges are obtained (e.g., examiners and good friends), or because they involve both global judgments and observations in specific kinds of situations (decreasing the consensus); self has better access to private thoughts and feelings or to a broader range of situations than others (increasing the variance of the self-descriptions).

Because numerous factors drive the self-enhancement correlation in either direction, they might cancel each other out to some extent. This argument seems to have some merits when self-enhancement in patterns of traits is studied (such as by Colvin et al., 1995, who studied the favorability of Q-sort patterns). When self-enhancement in specific traits is analyzed, however, it seems less applicable. Instead, a high variance of the self-enhancement correlation across different traits is expected because the previously mentioned factors are associated with different traits to different degrees.

Self-Enhancement Correlations and Trait Evaluativeness

In our empirical analysis, we used the variability of self-enhancement across different traits to test a fundamental assumption about self-enhancement: Interindividual differences in self-enhancement increase with increasing trait evaluativeness. Four hypotheses were derived from this assumption. Two of them concerned self-enhancement correlations: Self-enhancement correlations are zero for nonevaluative traits on average, and they become more negative with increasing trait evaluativeness. Both hypotheses were clearly confirmed by our analyses, both at the level of trait items and at the level of scales that were composed of items of a particular degree of evaluativeness.

The scale approach led to a stronger negative correlation between trait evaluativeness and self-enhancement correlations because it minimized the influence of factors other than self-enhancement on the self-enhancement correlations. The link between self-enhancement correlations and trait evaluativeness was stronger for trait variations in favorability than for trait variations in unfavorability. Thus, the enhancement of positive traits seems to be more closely associated with unfavorable traits than is the minimization of negative traits. From a motivational point of view, this difference suggests that need for approval is more strongly linked to unfavorable traits than avoid-
ance of disapproval, or that need for approval evokes more negative reactions from others than avoidance of disapproval (see Crowne & Marlowe, 1964).

The clear confirmation of these first two hypotheses lent further support to a substantive interpretation of self-enhancement correlations. Self-enhancement correlations vary around zero for nonevaluative traits because there is no reason to self-enhance one's self-description in these traits. Consequently, the interindividual variance in self-enhancement is close to zero and so is the self-enhancement correlation. The more favorable or unfavorable a trait is, the more individuals are differentially motivated to self-enhance their self-description in this trait. These interindividual differences, in turn, lead to a negative self-enhancement correlation. It is difficult to believe that these expected, systematic effects could be attributed to statistical artifacts.

**Consistency of Self-Enhancement Across Traits and Trait Evaluativeness**

The other two hypotheses concerned the consistency of self-enhancement across traits: Interindividual differences in self-enhancement are consistent across highly evaluative traits, and their cross-trait consistency increases with increasing trait evaluativeness. Both hypotheses were also clearly confirmed. The consistency of self-enhancement scores across highly evaluative trait items was so high that it suggested a rational scale construction for the measurement of interindividual differences in self-enhancement: Any highly evaluative items form a consistent self-enhancement scale, regardless of their content. Approximately 20 of such items suffice for a reliable scale.

**Self--Other Agreement, Consensus, and Trait Evaluativeness**

The self--other agreement was somewhat lower than the consensus of others although the difference was small at the level of trait items. Furthermore, this difference was related to trait evaluativeness, particularly for scales: The more a trait was evaluative, the lower the self--other agreement was relative to the consensus. These results replicated the results by John and Robins (1993), but with a much larger sample of traits. Both findings are expected from a self-enhancement perspective: Interindividual differences in self-enhancement decrease the self--other agreement, and this effect increases with increasing trait evaluativeness.

Our results were not completely consistent with those reported by John and Robins (1993), however. In their study, trait evaluativeness correlated negatively with both the self--other agreement and the others' consensus, whereas in the present study it correlated negatively with self--other agreement but positively with others' consensus at the item level, and positively with both kinds of agreement at the scale level. That the correlations for the scales were much more positive than for the items can be attributed to the fact that only the more evaluative scales contained shared item variance on which judges could consistently agree. It is not clear, however, why the consensus among others was differently related to the evaluativeness of individual items in our and John and Robins's (1993) studies. The difference may be due to more selective samples of traits in John and Robins's (1993) two studies (76 and 40 traits as compared with 830 trait items in our study).

**Variability of Self-Enhancement Across Traits**

The self-enhancement effects that we found in our large sample of traits varied considerably across traits. This cross-trait variability could be explained to some extent by differences in the evaluativeness of the traits, but this property of traits explained only a minor portion of the cross-trait variance at the level of items. The remaining variability can not be easily explained by measurement error due to lack of aggregation, because each self-enhancement correlation was based on approximately 500 participants and 1,000 acquaintances. If self-enhancement correlations vary considerably between traits, it seems useful to explore which traits show particularly negative, which traits show particularly positive, and which traits show no correlation with self-enhancement.

John and Robins (1994) found strong evidence that narcissism was positively related to self-enhancement in group discussions. Beyond this obvious relation (self-enhancement is an essential element of the definition of narcissism), the relation between other specific traits and self-enhancement is less clear. The correlational patterns reported by Colvin et al. (1995) convey no clear answer to this question. It can be explored in our larger set of trait items, but such an analysis is beyond the scope of this article.

**Average Versus Differential Self-Enhancement**

One unexpected finding merits some discussion. The participants in our sample showed a modest, overall self-devaluation effect: On average, they rated themselves somewhat higher on unfavorable traits and somewhat lower on favorable traits than their acquaintances. Viewed from the perspective of the acquaintances, they rated the participants more leniently than the participants rated themselves. Thus, the average self-enhancement effect was opposite to what one might infer from Taylor and Brown's (1988) study. Leniency effects are not unusual in studies of trait ratings. In Colvin et al.'s (1995) Study 1, for example, the mean favorability score for the self-description was .51, whereas it was .69 for the description by examiners and .54 for the description by friends.

It is important to emphasize that the main questions of this article concern differential self-enhancement, that is, interindividual differences in self-enhancement, not average self-enhancement. Apart from floor or ceiling effects caused by extreme average self-devaluation or self-enhancement, interindividual differences in self-enhancement and their correlates are independent of average self-enhancement. Thus, the question of whether self-enhancement correlations are positive or negative for a particular trait is independent of whether individuals self-enhance or self-devaluate on average.

**Resolving the Paradox: Taylor and Brown Versus Colvin, Block, and Funder**

The distinction between average self-enhancement and differential self-enhancement is also useful for understanding a subtle
but quite consequential difference between the approaches by Taylor and Brown (1988) and Colvin et al. (1995) who apparently came to different conclusions about the appropriateness of self-enhancement. Taylor and Brown (1988) reported that people characterized by low self-esteem or mild depression self-devaluate more than people on average. From this finding, they concluded that self-enhancement is healthy. Colvin et al. (1995) found that interindividual differences in self-enhancement are negatively related to favorable personality traits in unselected samples. From this finding they concluded that self-enhancement is unhealthy. These two conclusions are not incompatible: Both effects can co-exist in the same population because Taylor and Brown (1988) were concerned only with a small subgroup of unfavorable personalities (see Figure 6).

Colvin et al. (1995) were interested in the correlation between mental health (favorable personality) and self-enhancement in unselected samples (such a sample is visualized by the ellipse in Figure 6). Their finding of a negative correlation is graphically displayed in Figure 6 as a regression line with a negative slope. Taylor and Brown (1988) contrasted the average self-enhancement in such samples (the large cross) with the average self-enhancement in a special sample of mentally unhealthy people (the small cross; the special sample is marked by the small circle). Their finding that unhealthy people enhance themselves less than an unselected sample is graphically displayed in Figure 6 by a regression line with a positive slope. If this special sample contributes little to the overall correlation studied by Colvin et al. (1995), Taylor and Brown's and Colvin et al.'s results would be compatible: They can co-exist in the same population.

Thus, the heart of the problem is that Taylor and Brown inferred from the low self-enhancement of a particular mentally unhealthy group an overall differential effect for the whole population ("self-enhancement fosters mental health"). Such an inference can be misleading if other mentally unhealthy groups exist in the population that do not show the self-enhancement tendency that Taylor and Brown (1988) had observed for people with low self-esteem or mild depression. This becomes more obvious when we consider how mental health was operationalized in these studies.

The Many Faces of "Mental Health"

Taylor and Brown (1988) referred to mental health problems that are related to dependent, avoidant, and depressive personality disorders as defined in the Diagnostic and Statistical Manual of Mental Disorders (DSM-III, DSM-IV; American Psychiatric Association, 1987, 1994). Another class of mental health problems is described in DSM-III and DSM-IV as antisocial and narcissistic disorders. In a conjoint factor analysis of personality disorder scales, Wiggins and Pincus (1994) found that the first factor that accounted for most variance contrasted dependent–avoidant disorders with antisocial–narcissistic disorders. It is not unreasonable to assume that people with antisocial or narcissistic disorders display above-average self-enhancement (see John & Robins, 1994). If Taylor and Brown (1988) had contrasted antisocial or narcissistic personalities with people on average, they might have come to the same conclusion as Colvin et al. (1995): Self-enhancement is unhealthy.

Thus, different mental health problems may be related to self-enhancement in very different ways. Nonlinear relationships are also possible between mental health and self-enhancement. For example, both extreme self-enhancement and extreme self-devaluation may indicate different mental health problems whereas moderate self-enhancement may characterize the healthy personality (see Block & Thomas, 1955, for a similar argument).

Colvin et al.'s (1995) approach included both kinds of problems because both are judged as unfavorable. Therefore, their approach is more appropriate for the general question of whether self-enhancement is healthy. But if different kinds of mental health problems affect self-enhancement differently, neither Taylor and Brown's (1988) nor Colvin et al.'s (1995) nor our approach are sufficient. Studies are needed that compare self-enhancement across different kinds of mental health problems, either from a type perspective (e.g., personality disorders as assessed through DSM-IV), or from a dimensional perspective (e.g., antisocial or narcissistic tendency versus dependent, avoidant, or depressive tendency). Such more differentiated approaches might lead to a more differentiated view of the role of self-enhancement in mental health.

References


IS SELF-ENHANCEMENT HEALTHY?

Appendix

Statistical Properties of the Self-Enhancement Correlation

Let $X$ denote a self-assessed trait (e.g., a self-rating or the favorability of a Q-sort self-description), and $Y_1$, $Y_2$ assessments of the same trait by others. Without limiting generality, we can assume that the variance of $Y_2$ equals 1 (if not, we simply divide $X$, $Y_1$, $Y_2$ by the standard deviation of $Y_2$). We do not make any assumptions about the distributions of $X$ and $Y_1$. These variables can be decomposed as follows:

$$X = T + X_1 + X_2 + E,$$

with $r(T, X_1) = r(T, X_2) = r(T, E) = 0,$ \hspace{1cm} (A1)

where $T$ represents the "true" interindividual variance that is shared by self and both others, $X_1$ (or $X_2$) represents the additional variance that is shared by self and other 1 (or 2), and $E$ represents the measurement error of self, and $r$ denotes the Pearson correlation.

$$Y_i = T + O + X_i + E_i,$$

with $r(T, O) = r(T, X_i) = r(T, E_i)$ = $r(O, X_i) = r(O, E_i)$ = $r(X_i, E_i) = 0$, \hspace{1cm} (A2)

for $i = 1, 2$, where $O$ represents the variance that is shared by both others, but not by self, and $E_i$ represents the $i$th other's measurement error.

Thus, all component variables are assumed to be linearly independent. Under these assumptions, it follows:

$$s(X) \cdot r(X, Y_1) = \text{cov}(T + X_1 + X_2 + E, T + O + X_1 + E_1),$$

where $s$ denotes the standard deviation and $\text{cov}$ denotes the covariance

$$= \text{cov}(T, T) + \text{cov}(X_i, X_j),$$

$$s(X) \cdot r(X, Y_2) = s^2(T) + s^2(X_i);$$

$s(X_1) \cdot r(Y_1, Y_2) = \text{cov}(T + X_1 + X_2 + E, T + O + X_1 + E_1)$

$$= \text{cov}(T, T) + \text{cov}(X_i, X_j),$$

$s(Y_1) \cdot r(Y_1, Y_2) = \text{cov}(T + O + X_i + E_i, T + O + X_i + E_2)$

$$= \text{cov}(T, T) + s^2(T) = s^2(T) + s^2(O).$$. (A3)

Thus,

$$s(X) \cdot r(X, Y_2) - s(Y_i) \cdot r(Y_1, Y_2) = s^2(X_2) - s^2(O).$$ (A4)

Using the linear independence of the components in (1), (2), the "self-enhancement correlation" $r(X - Y_1, Y_2)$ can be expressed as shown in Figure A1. Because of Equations A6 and A3, it follows

$$r(X - Y_1, Y_2) = \frac{s(X) \cdot r(X, Y_2) - s(Y_i) \cdot r(Y_1, Y_2)}{\sqrt{s^2(X) + s^2(Y_1) - 2 \cdot s(X) \cdot s(Y_1) \cdot r(X, Y_1)}}.$$ (A5)

If $s(X) = s(Y_1)$, it follows

$$r(X - Y_1, Y_2) = \frac{s(X) \cdot [r(X, Y_2) - r(Y_1, Y_2)]}{\sqrt{s^2(X) \cdot [1 - r(Y_1, Y_2)]}}.$$ (A6)

If in addition $Y_1 = Y_2$, the formula for the ordinary regression to the mean follows:

$$r(X - Y, Y) = \frac{[r(X, Y) - 1]}{\sqrt{2 \cdot [1 - r(X, Y)]}}.$$ (A7)

Because $Y_1$, $Y_2$ serve exchangeable functions in Equation A7, it is reasonable to assume:

(a) $s(Y_i) = s(Y_2) = 1$,

(b) $r(X, Y_1) = r(X, Y_2)$.

Under these two assumptions, we studied self-enhancement correlations as a function of the difference between self--other agreement and consensus, and of the standard deviation of the self-assessment at four levels of self--other agreement. Figure A2 shows the resulting functions.

Figure A2 indicates that self-enhancement correlations increase linearly whenever (a) the self--other agreement decreases relative to the consensus among others, and (b) the standard deviation of the self-assessment decreases relative to the standard deviation of the other- assessments; the latter effect becomes stronger with increasing self--other agreement.

$$r(X - Y_1, Y_2) = r((T + X_1 + X_2 + E) - (T + O + X_1 + E_1), Y_2)$$

$$= r(X_2 + E - O - E_1, T + O + X_2 + E_2)$$

$$= [\text{cov}(X_2, X_2) - \text{cov}(O, O)]/\sqrt{s^2(X_2) + s^2(E) + s^2(O) + s^2(E)}$$

$$= [s^2(X_2) - s^2(O)]/\sqrt{s^2(X) - s^2(T) - s^2(X_1) + s^2(Y_1) - s^2(T) - s^2(X_1)}$$

$$= [s^2(X_2) - s^2(O)]/\sqrt{s^2(X) + s^2(Y_1) - 2 \cdot s^2(T) + s^2(X_1)}.$$

Figure A1. Equation for the self-enhancement correlation.

(Appendix continues)
Figure A2. Self-enhancement correlations as a function of their components. - - - - \( s(\text{self})/s(\text{other}) = 1.5 \); \( s(\text{self})/s(\text{other}) = 1.25 \); \( s(\text{self})/s(\text{other}) = 1.00 \); \( s(\text{self})/s(\text{other}) = 0.75 \); \( s(\text{self})/s(\text{other}) = 0.5 \).