Running Head: INTROSPECTION AND SELF-ENHANCEMENT

The Why's the Limit:

Curtailing Self-Enhancement with Explanatory Introspection

Constantine Sedikides
University of Southampton

Robert S. Horton
Wabash College

Aiden P. Gregg
University of Southampton

Abstract

Self-enhancement is linked to psychological gains (e.g., subjective well-being, persistence in adversity), but also to intrapersonal and interpersonal costs (e.g., excessive risk-taking, antisocial behavior). Thus, constraints on self-enhancement may sometimes afford intrapersonal and interpersonal advantages. We tested whether explanatory introspection (i.e., generating reasons for why one might or might not possess personality traits) constitutes one such constraint. Experiment 1 demonstrated that explanatory introspection curtails self-enhancement. Experiment 2 clarified that the underlying mechanism must (a) involve explanatory questioning rather than descriptive imagining, (b) invoke the self rather than another person, and (c) feature written expression rather than unaided contemplation. Finally, Experiment 3 obtained evidence that an increase in uncertainty about oneself mediates the effect.

The Why's the Limit:
Curtailing Self-enhancement with Explanatory Introspection

Most people, most of the time, see themselves through rose-colored glasses. Whether rating themselves as above-average on personality traits and abilities (Alicke, 1985) or believing themselves less susceptible to bias than the average person (Pronin, Yin, & Ross, 2002)—whether showing selective recall for flattering autobiographical episodes (Sanitioso, Kunda, & Fong, 1990) or engaging in social comparisons that validate a positive self-view (Dunning, 1999)—whether attributing their successes internally and their failures externally (Mezulis, Abramson, Hyde, & Hankin, 2004) or thinking that their own future will surpass that of their peers (Weinstein, 1980)—people by and large evaluate themselves more favorably either than the objective facts warrant (Gosling, John, Craik, & Robins, 1998) or than external observers think justified (Epley & Dunning, 2000). Tellingly, people even believe that they outdo their own doppelgangers: they rate themselves more favorably than they rate their peers on the basis of identical behavioral evidence (Alicke, Vredenburg, Hiatt, & Govorun, 2001). Moreover, egocentric biases like the better-than-average effect are pervasive existing not only in (self-promoting) individualistic cultures, but also in (self-deprecating) collectivistic cultures (Sedikides, Gaertner, & Toguchi, 2003).

All such phenomena can be viewed as forms of self-enhancement. Although perhaps irrational in the normative sense—half of us being forever doomed to be below average[1]—self-enhancement is nonetheless linked to substantial benefits. These include good psychological health (Taylor, Lerner, Sherman, Sage, & McDowell, 2003), better coping with physical illness (Taylor et al., 2003) and traumatic loss (Bonanno, Rennicke, & Dekel, 2005), greater persistence in the face of adversity (Taylor & Brown, 1988), and good social adjustment (Donnellan, Trzesniewski, Robins, Moffitt, & Caspi, 2005).

However, self-enhancement is also linked to several substantial costs. Intrapersonal costs include imprudent risk-taking (Baumeister, Heatherton, & Tice, 1993), ineffective action planning (Oettingen & Gollwitzer, 2001), and an increased likelihood of disengaging from academic studies (Robins & Beer, 2001). Interpersonal costs involve being perceived negatively and treated unpleasantly by others. For example, after a brief period of infatuation, peers come to regard inveterate self-enhancers as conceited, defensive and hostile (Paulhus, 1998), and are generally prone to deride them, if not isolate them interpersonally (Schlenker & Leary, 1982). In addition, concerns about promoting or protecting a favorable public self-image can prompt actions that lead to illness, injury, and death: Notoriously, people from temperate climes often sunbathe for hours to look and feel good among their peers, thereby raising their risk of sunstroke, sunburn, and skin cancer (Leary, Tchividjian, & Kraxberger, 1994).

In view of these inauspicious correlates, it is perhaps salutary that self-enhancement, although pervasive, is not inevitable: it varies naturally and can be strategically manipulated. For example, the topic of judgment moderates self-enhancement: people self-enhance less on traits that lack ambiguity (Dunning, Meyerowitz, & Holzberg, 1989) or that they believe they can modify (Dauenheimer, Stahlberg, Spreeman, & Sedikides, 2002). In addition, several interpersonal factors are also known to constrain self-enhancement. These include the similarity of the comparison other to the self (Stapel & Schwinghammer, 2004), the concreteness of the comparison other (Alicke, Klotz, Breitenbecher, Yurak, & Vredenburg, 1995), concerns about preserving close relationships (Tice, Butler, Muraven, & Stillwell, 1995), and social pressures to be accountable (Sedikides, Herbst, Hardin, & Dardis, 2002). However, given the problems that self-enhancement sometimes poses, it is worth exploring what other factors have the potential to curtail it. In this article, we investigate a possible intrapersonal factor: introspection.

Varieties of Introspection

The human ability to introspect has long fascinated philosophers. Descartes (see Cottingham, Stoothoff, & Murdoch, 1984) regarded reflexive thought as proof of an indubitable self. Introspection has also captivated the attention of psychologists, from the early structuralists (Titchener, 1912; Wundt, 1894) to modern-day experimental social psychologists (Hirt & Markman, 1995; Hixon & Swann, 1993; Wilson, Dunn, Kraft, & Lisle, 1989). Importantly, introspection is considered a uniquely human capacity (Sedikides & Skowronski, 1997, 2000;

Sedikides, Skowronski, & Dunbar, 2006) and its investigation is central to personality and social psychology (Bless & Forgas, 2000; Maio & Olson, 1998; Wilson & Dunn, 2004).

Conceptual Distinctions

Introspection is the process of looking inward, thinking “about [one’s] thoughts and feelings” (Wilson et al., 1993, p. 33), or about oneself as a whole. However, introspection is not a unitary construct. Indeed, it can be conceptualized in at least two distinct ways.

One type of introspection constitutes what we term *descriptive introspection*. This denotes the act of contemplating what one’s personality is like. When introspecting descriptively, people ask themselves questions like “Do I have (or not have) traits X and Y?” or “To what extent do I have (or not have) traits X and Y?” People then conclude that they possess or lack particular traits to some degree or other. Another type of introspection constitutes what we term *explanatory introspection*. This denotes the act of contemplating *why* one does or does not think of oneself in a particular way. When introspecting explanatorily, people ask themselves questions like “Why might I have (or not have) traits X and Y?” or “What are the reasons for my having (or not having) traits X and Y?” People then generate reasons that explain why they either possess or lack particular traits to some degree or other.

Descriptive and Explanatory Introspection: A Review of the Literature

Descriptive and explanatory introspection, or key elements thereof, have already been operationalized as independent variables in past research. Consider two lines of inquiry. First, Tesser (1978) investigated the consequences of thinking about an attitude object for which a well-developed knowledge base exists. Intensive thinking led to the formation of an evaluatively-consistent belief set, which in turn polarized attitudinal judgments. That is, intensive thinking produced “more univalent, less ambivalent” attitudes (p. 295). Second, Hixon and Swann (1993: Experiment 3) had participants peruse particular dimensions of personality. In particular, undergraduates with low self-esteem pondered the question “What kind of person are you in terms of sociability, likeability, and interestingness?” while weighing up the accuracy of two evaluations—one flattering and one critical—that graduate students ostensibly provided of them. Consistent with their pre-existing negative self-view, the undergraduates endorsed the critical evaluation over the flattering one.

The two lines of inquiry have common elements. First, in terms of procedure, participants either reviewed a stored body of knowledge, or answered a “what” question. Both activities are clearly reminiscent of descriptive introspection. Second, in terms of outcome, participants either consolidated an attitude or confirmed a self-view. Either way, a previously held belief was strengthened. The conjunction of these facts suggests that descriptive introspection is a source of psychological stability (Silvia & Gendolla, 2001).

In other lines of research, examples of explanatory introspection are clearly discernible. Wilson and his colleagues have investigated the impact of this type of introspection on attitudes towards various objects (e.g., the self, political candidates, collegiate classes; Wilson, Dunn et al., 1989). Participants wrote down some reasons why they liked or disliked an object and thereafter expressed their attitudes toward that object. Reasons-analysis perturbed attitudes, prompting either a shift in direction or an increase in variability (Wilson et al., 1993). This perturbation was attributed to the temporary accessibility of reasons that, although easily verbalized and subjectively plausible, are nonetheless unrepresentative of the full set of reasons and at odds with dispositional preferences. Similar experimental procedures, findings and explanations apply to a line of research on value change by Maio, Olson, and colleagues (Bernard, Maio, & Olson, 2003a; Maio & Olson, 1998; Maio, Olson, Allen, & Bernard, 2001). Both research programs suggest that explanatory introspection is an agent of psychological change.

Explanatory introspection also features in research on explanatory bias. Participants, when instructed to explain why a particular hypothetical outcome might occur, overestimate the likelihood of its occurrence (Ross, Lepper, Strack, & Steinmetz, 1977). The bias is observed regardless of whether the to-be-explained outcome pertains to the self (Kunda & Sanitioso, 1989), to another person (Anderson, 1982), or to an event like a political election (Caroll, 1978) or sporting competition (Markman & Hirt, 2002). As in the attitudes/values literature, information availability and accessibility have been invoked as underlying mechanisms. In particular, the goal of explaining some outcome prompts an information search that brings outcome-consistent arguments to the forefront of the mind (Tversky & Kahneman, 1974), where they influence, in an assimilative manner, the ensuing judgment (Kunda & Sanitioso, 1989). An alternative account of explanatory bias posits that the goal of outcome explanation prompts a frame of mind in which the explanation (or focal hypothesis; Koehler, 1991) is assumed to be true. Evidence is then reviewed from the perspective of that frame, and thus selectively accumulates in the direction of the focal hypothesis, leading its merits to be overestimated (Hirt & Markman, 1995). Regardless of the underlying mechanism, research on explanatory bias suggests that explanatory introspection has well-defined directional effects.

Finally, explanatory introspection features in debiasing research. In a typical task, participants are presented with an event and instructed to explain how it might give rise both to one outcome and to another (alternative or contrary) outcome. This task—variants of which are known as counterexplanation, consider-the-opposite, inoculation, or consider-an-alternative—attenuates the magnitude of the explanatory bias (Anderson, 1982; Hirt & Markman, 1995; Hirt, Kardes, & Markman, 2004; Lord, Lepper, & Preston, 1984). This body of research suggests that explanatory introspection, when it involves a consideration of more than one point of view, exerts a moderating influence on psychological processes.

Taking our cue from the above lines of research, we wondered whether explanatory introspection could curtail self-enhancement. We accordingly devised an introspection manipulation that blended elements of a prototypical debiasing manipulation with elements of a typical reasons-analysis manipulation. Specifically, we had participants generate reasons for why they might or might not have a set of important personality traits.[2] Two key features of our adaptation are worth noting. First, our participants focused on the self rather than on a hypothetical person, object, or event. Second, our participants focused on central (or core) facets of the self (Sedikides, 1993). Thus, with the self involved, our particular adaptation likely facilitated the emergence of motivational processes above and beyond conventional cognitive ones. Any account of underlying mechanisms would need to take this into consideration.

Pretesting

First off, we ran a pretest in order to identify a set of nomothetic trait dimensions that participants would regard as central to their self-concept. In this pretest—as in all subsequently reported experiments—participants were undergraduates from the University of North Carolina at Chapel Hill, fulfilling an introductory psychology course option.

Central trait dimensions can be operationally defined as those that elicit extreme ratings when it comes to three pertinent properties: self-descriptiveness (i.e., either highly self-descriptive or not at all self-descriptive), valence (i.e., either highly positive or highly negative), and importance (i.e., very important to have or very important not to have). Sixty-five participants duly rated 24 trait adjectives—corresponding to the positive and negative poles of 12 trait dimensions—in terms of all three properties (Table 1). Central trait dimensions were then selected for use, if two conditions were met. First, the positive pole of the dimension had to be rated among the four most self-descriptive, most positive, and most important to have; second, the negative pole of the dimension had to be rated among the four least self-descriptive, least positive (i.e., most negative), and least important to have. These selection criteria yielded three central trait dimensions: honest-dishonest, kind-unkind, and trustworthy-untrustworthy. These trait dimensions were subsequently broken down into two contrasting categories of trait adjective for use in the experiments: central positive (honest, kind, trustworthy), and central negative (dishonest, unkind, untrustworthy).

Experiment 1

The objective of Experiment 1 was to test whether explanatory introspection curtails self-enhancement. We instructed participants to analyze the reasons both for why they might and might not have a particular trait. Additionally, we asked some participants to introspect explanatorily about positive traits, others about negative traits. Participants in the control group engaged in a neutral task irrelevant to self. Our prediction was that, compared to control participants, explanatory introspection participants would self-enhance less by giving both lower self-ratings on positive traits and higher self-ratings on negative traits.

Method

Participants and Experimental Design

Eighty-eight participants were randomly assigned to a 2 (Cognitive Activity: Explanatory Introspection vs. Control) X 2 (Trait Valence: Positive vs. Negative) balanced factorial design. In this and all subsequent experiments, participants were tested individually and debriefed thoroughly.

Procedure

Participants assigned to the two Explanatory Introspection cells were instructed to generate reasons for why they might or might not have each of three traits. In the Positive cell, the traits in question were honest, kind and trustworthy, and in the Negative cell, dishonest, unkind and untrustworthy. The instructions read as follows:

> “We are interested in the *reasons why* you might or might not have the trait ___Please take a few moments to think about *why* you might or might not have the trait ___.We want you to analyze *very carefully* the reasons you might or might not have the trait ___ because this will help you organize your thoughts for subsequent tasks.”

Participants were encouraged one final time to analyze very carefully why they both might and might not have each trait, and were then asked to write the relevant reasons down, using a separate page for each trait. Participants assigned to the two Control cells instead listed as many uses as possible—again, positive or negative, depending on the cell— for three everyday objects (spoon, brick, and briefcase; cf. Sedikides, Campbell, Reeder, & Elliot, 1998), and again used a separate page for each item. Participants were told that all pages were theirs to keep if they so desired so as to encourage frank responding. However, all opted to leave the pages behind in the experimental booth.

Next, all participants (including controls) rated the self-descriptiveness of three traits (positive or negative, depending on the experimental condition). In particular, they responded to the question: “To what extent do you think you have the trait ___?” (1 = *not at all*, 15 = *very much*). Finally, to explore underlying mechanism, Explanatory Introspection participants (but not Controls) labeled each reason that they generated as either “confirming” or “disconfirming” the trait they had considered.

Results

Self-Evaluation

The three trait self-descriptiveness ratings were internally consistent ($\alpha = .95$) and so averaged to form a composite index. We then entered this index to a two-way factorial ANOVA (Cognitive Activity X Trait Valence). A significant main effect for Trait Valence emerged: Participants rated positive traits ($M = 12.64$) as more self-descriptive than negative traits ($M = 3.36$), $F(1, 84) = 1402$, $p < .001$, replicating a well-established finding (Sedikides, 1993).

More importantly, this main effect was qualified by a predicted interaction, $F(1, 84) = 9.67$, $p < .005$. Explanatory introspection participants regarded positive traits ($M = 12.20$, $SD = 1.38$) as significantly less self-descriptive than controls did ($M = 13.09$, $SD = .98$), $F(1, 42) = 6.10$, $p < .02$, and regarded negative traits ($M = 3.68$, $SD = 1.35$) as marginally more self-descriptive than controls did ($M = 3.03$, $SD = .85$), $F(1, 42) = 3.66$, $p < .06$. That is, explanatory introspection participants, compared to controls, evaluated themselves less positively and (tendentially) more negatively. In sum, explanatory introspection curtailed self-enhancement.

Reasons Generated

On the basis of past research, we expected that, during explanatory introspection, participants would engage in autobiographical searches, retrieving episodic or abstract information from long-term memory. This was indeed the case. In this and subsequent experiments, the reasons that participants gave (a) were non-overlapping, and (b) consisted almost uniquely of episodic memories or habitual behaviors, for example, “I [once] lied to parents about where I went at night” (confirming dishonest) and “I [typically] tell people actually what I think about them” (confirming honest).

We also expected that the reasons participants generated would correspond intelligibly to their self-descriptiveness ratings. To begin with, we summed the total number of reasons that each participant labeled as confirming a trait, and then divided this by the total number of reasons they generated for that trait. We derived such a ratio separately for each trait, and then created a composite confirmation index by averaging all three ratios ($\alpha = .75$). A one-way ANOVA incorporating this index showed that participants generated a significantly higher proportion of confirming reasons when they considered positive traits ($M = .73$) than when they considering negative ones ($M = .38$), $F(1, 42) = 35.57$, $p < .001$. Interpreted somewhat differently, participants confirmed their positive but disconfirmed their negative traits, replicating past research (Dunning et al., 1989; Sedikides, 1993).

More importantly, we investigated whether the confirmation index correlated significantly with participants' self-descriptiveness scores. It did, $r(42) = .72$, $p < .001$. This result suggests that explanatory introspection participants based their self-descriptiveness ratings largely on the reasons they generated. Moreover, this account is in keeping with previous research showing that the generation of supportive thoughts increases the endorsement of personality characteristics (Davies, 2003). However, alternative accounts—for example, that reasons were based on self-descriptions—cannot be definitively ruled out. (We investigate the matter further in Experiment 3.) Note that the correlations between the confirmation index and self-descriptiveness scores for participants considering positive traits ($r[20] = .27$, $p < .23$) and negative traits ($r[20] = .40$, $p < .07$) did not differ significantly from one another, $z = .45$, $p < .65$.

Summary

Relative to Controls, participants who explanatorily introspected showed an attenuated tendency to self-enhance. In particular, they regarded positive traits as significantly less self-descriptive, and negative traits as marginally more self-descriptive. Regardless of trait valence, self-descriptiveness scores correlated with confirmatory reasons generated via explanatory introspection, suggesting that self-judgments varied as a function of the accessibility of autobiographical instances.

Discussion

What are the psychological mechanisms by which explanatory introspection curtails self-evaluation? Explanatory introspection both reduced the positivity of self-views on positive dimensions and tended to increase the negativity of self-views on negative dimensions. Any comprehensive account must therefore explain why self-enhancement was attenuated in both cases.

The first point to note is that, given the ubiquity and common pre-eminence of the self-enhancement motive (Baumeister, 1998; Sedikides & Gregg, 2003), participants' levels of self-regard were likely approaching their upper limit. This is because, to the extent that people can self-enhance, they generally will: the balloon of self-regard will rise as far as the ballast of rational and normative constraints permits (Sedikides & Strube, 1997). At the start of the experimental session, our participants, already fairly high-achieving members of a Western culture, would not have been under any special pressure to self-derogate. Their levels of self-regard would likely have been closer to their maximum than their minimum. Thus, their self-regard would have had more room for maneuver in a downward direction than in an upward one, regardless of whether they explanatorily introspected about positive traits or about negative ones. Hence, any factor undermining self-regard would have observably reduced it more than any intrinsically comparable factor promoting self-regard would have observably increased it.

The second point to note is that, generally speaking, negative factors exert a greater impact than positive ones (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001). To take one of numberless examples, the prospect of losing a substantial sum of money strikes most people as more aversive than the prospect of gaining that sum strikes them as attractive (Kahneman & Tversky, 1981). Now, explanatory introspection participants were instructed to consider, not only why they might possess, but also why they might not possess, particular traits. Thus, when those traits were positive, participants considered both why they might possess them (an attractive reflection) and why they might not (an aversive reflection); and when those traits were negative, participants considered both why they might possess them (an aversive reflection) and why they might not (an attractive reflection). Given the generally greater power of negative factors, it would hardly be surprising if participants' aversive reflections exerted greater psychological impact than the attractive reflections. If they did—and if, as seemed to have been the case, their self-regard varied as a function of the reasons they generated—then the net result would have been a reduction in self-enhancement.

The combination of both dynamics plausibly accounts in general for why explanatory introspection curtails self-enhancement, regardless of whether positive or negative traits are considered. Of course, this is only a distal outline; the proximal details still require filling in. The effects of explanatory introspection are likely proximally mediated by induced variations in the accessibility of self-knowledge (Davies, 2003; Fazio, Effrein, & Falender, 1981; Schwarz et al., 1991). Explanatorily introspecting participants, when attempting to answer self-generated questions about whether they possess or lack personality traits, will engage in retrospective mental simulations (Sanna, 2000) and autobiographical memory searches (Kihlstrom, Beer, & Klein, 2003). Such simulations and searches will prompt consideration of a relatively broad set of plausible alternatives. Participants will bring to mind both instances in which they behaved in a trait-confirming manner and instances in which they behaved in trait-disconfirming manner. The relative accessibility of these instances, accompanied by a state of heightened self-uncertainty (Petty, Brinol, & Tormala, 2002), will then trigger corresponding self-judgments (i.e., trait self-descriptiveness ratings). In terms of the two dynamics discussed above, negatively-toned simulations and searches are liable to be rendered more accessible, or to be weighted more heavily, than positively-toned ones; and, given the normative positivity of self-regard, such negatively-toned simulations will have greater scope for impact.

Experiment 2

One purpose of Experiment 2 was simply to replicate Experiment 1. We therefore included experimental and control conditions permitting the effects of explanatory introspection to be tested, both when positive and negative traits were considered. But Experiment 2 had an additional purpose: to pin down the precise preconditions for curtailing self-enhancement through explanatory introspection. This necessitated some methodological additions and theoretical extensions.

First, we wondered whether the active ingredient of our manipulation might be the more general act of asking explanatory questions about personality traits (or anything else) rather than the more specific act of asking explanatory questions about one's own personality traits. Do inquiries have to be self-directed in order for self-enhancement to be curtailed, or will other-directed inquiries suffice? Because only self-directed inquiries constitute introspection, this question needed to be addressed. To address it, we directly manipulated the target of scrutiny (Target Type). In particular, we had half the participants consider their own personality traits (Self), and the other half an acquaintance's personality traits (Other). We predicted that self-enhancement would be curtailed only in the Self condition. Note that this distinction between self-directed and other-directed inquiry parallels one drawn by previous researchers (Klein & Loftus, 1988; Sedikides & Green, 2000), who argued that different cognitive processes are at work when individuals process self-related versus other-related information: elaboration in the first case (i.e., considering a new instance in relation to prior self-knowledge), organization in the second (i.e., considering a new instance in relation to other instances).

Second, we further explored the hypothesis that temporary self-knowledge accessibility mediates the impact of explanatory introspection on self-enhancement. As before, we asked all participants in the Explanatory Introspection condition to generate reasons why they might have or not have a set of traits. This time, however, we instructed only half of them to list those reasons in written form, and instructed the other half merely to entertain those reasons in mental form. We labeled this variable Activity Type (Written vs. Mental). We suspected that the requirement to write reasons down would be a critical factor in success of the manipulation. For one thing, the act of writing something down is liable to concretize and stabilize thoughts that would otherwise remain hypothetical and fleeting; this, in turn, is liable to increase durably the accessibility of trait-related thoughts and their derivative associations (cf. Pennebaker, 2003). For another thing, the act of writing something down is liable to engender consistency motivation by committing participants to the content of statements willingly expressed (Festinger & Carlsmith, 1959) or increasing a sense of accountability (Tetlock, Skitka, & Boettger, 1989); this, in turn, is liable to increase durably the weight ascribed to the underlying thoughts and associations. Hence, we predicted that the effects of explanatory introspection would be present in the Written condition but not in the Mental condition.

Third, past research suggests that, whereas explanatory introspection instigates a relatively impartial search of relevant autobiographical details (i.e., one that promotes psychological change), descriptive introspection instigates a relatively biased search (i.e., one that preserves psychological consistency; Tesser, 1978; Hixon & Swann, 1983). Hence, only explanatory introspection should curtail self-enhancement: descriptive introspection should merely maintain it. We tested this hypothesis by manipulating Inquiry Type (Explanatory vs. Descriptive). In particular, half of the participants considered the reasons why they (or someone else) did or did not possess particular traits (Explanatory), whereas the other half merely considered the extent to which they (or someone else) did or did not possess particular traits (Descriptive). (*Note*: In the Mental condition, Descriptive participants thought about the extent of trait possession, whereas in the Written condition, they committed those thoughts to paper.) We predicted that only explanatory participants (inquiring about self) would show moderation of self-regard on positive traits and extremification of self-regard on negative traits.

Finally, we modified our key manipulation slightly to reinforce its construct validity. In Experiment 1, both explanatory and control participants were free to take as much time as they needed to complete the task at hand. This methodological imperfection left the door open for possible confounds. For example, explanatory participants may have taken longer than control participants. If so, then the findings of Experiment 1 may simply have been due to more protracted cognitive activity. Hence, we standardized the task completion time to eliminate such temporal confounds. Specifically, all participants were allotted three minutes per trait.

In summary, Experiment 2 tested the boundary conditions of the self-enhancement curtailment effect observed in Experiment 1. We predicted that this effect would be observed only (or primarily) when participants (a) engaged in self-directed inquiries (as opposed to other-directed ones), (b) listed relevant considerations in writing (as opposed to merely mentally entertaining them), and (c) engaged in explanatory (as opposed to descriptive) introspection.

Method

Participants and Experimental Design

One-hundred and sixty participants were randomly assigned to one of 16 experimental conditions yielded by a 2 (Target Type: Self vs. Other) X 2 (Activity Type: Written vs. Mental) X 2 (Cognitive Activity: Explanatory vs. Descriptive) X 2 (Trait Valence: Positive vs. Negative) balanced factorial design.

A further 20 participants were randomly assigned to one of two control conditions (Trait Valence: Positive vs. Negative) identical to those in Experiment 1. The purpose of these control conditions was to test the replicability of Experiment 1, and to permit an additional test of the hypotheses of Experiment 2.

Procedure

Participants in the Explanatory condition were instructed to generate reasons (in written or mental form) for why someone (either they or another person) might or might not have three traits (either positive or negative). Instructions and traits dovetailed those of Experiment 1. Participants in the Descriptive condition were instructed to describe the extent to which someone might or might not have each trait.

Participants in the Self condition directed their trait-related inquires towards themselves, whereas those in the Other condition directed their trait-related inquiries towards an acquaintance. Before beginning, the latter wrote down the name of an acquaintance, and then stated (a) how many times they had interacted with him or her, (b) how well they knew him or her, and (c) how positive or negative their impression of him or her was. On average, participants reported that they had interacted with the acquaintance several times (M = 5.36 times) but they did not (yet) know him or her very well (M = 3.69, on a 9-point scale ranging from 1 = *not well* *at all* to 9 = *very well*), although they had nonetheless formed a mildly positive impression of him or her ($M = 6.24$, on a 9-point scale ranging from 1 = *very negative* to 9 = *very positive*).

Participants in the Written condition were instructed to list, on a separate sheet for each trait, the reasons (or thoughts) they had generated. Participants in the Mental condition were instructed that they need not to write anything down: it would suffice to generate the relevant reasons (or thoughts) in their head.

After being asked to generate reasons why (or thoughts about the extent to which) they might and might not possess each trait, all experimental participants were informed that they could generate as many or as few reasons (or thoughts) as they wished, but that they must do so within three minutes. Participants in the Control condition, working to the same deadline, were instructed to list as many uses as possible for a spoon, brick, and briefcase. All but 11 participants opted to leave the reasons pages behind in the experimental booth.

The final manipulated factor, Trait Valence, applied to both experimental and control participants. In different conditions, the former considered either three positive or three negative traits, and the latter either positive or negative uses for three objects. Finally, all participants completed self-descriptiveness trait ratings, as they had in Experiment 1.

Results and Discussion

Self-Evaluation

Being internally consistent ($\alpha = .94$), the three trait self-descriptiveness ratings were again averaged to form a composite index. We then entered this index into a four-way factorial ANOVA (Target Type X Cognitive Activity X Trait Valence X Activity Type). Replicating Experiment 1, a significant main effect for Trait Valence emerged, with participants endorsing positive traits ($M = 12.21$) more strongly than negative traits ($M = 3.92$), $F(1, 144) = 1164$, $p < .001$.

Importantly, this main effect was qualified by a three-way interaction between Target Type, Cognitive Activity, and Trait Valence, $F(1, 144) = 3.96$, $p < .05$. To clarify its meaning, we then examined the two-way Cognitive Activity X Trait Valence interaction separately for each level of Target Type (Other vs. Self). For Other, the interaction was not significant, *F*(1,72) < 1; for Self, it was, *F*(1, 72) = 4.11, *p* < .05. Specifically, Explanatory participants in the self condition endorsed positive traits marginally less strongly than Descriptive participants (*Ms* = 11.68 vs. 12.63), *F*(1, 36) = 3.44, p < .07; they also endorsed negative traits nonsignificantly more strongly (*Ms* = 4.62 vs. 3.92), *F*(1, 36) = 1.22, *p* < .28. This suggests that, averaging across Activity Type, explanatory introspection curtailed self-enhancement overall (relative to descriptive introspection).

However, the above three-way interaction was in turn qualified by Activity Type, to yield the predicted four-way interaction, *F*(1, 144) = 4.67, *p* < .04 (Table 2). We decomposed it by examining the three-way Cognitive Activity X Trait Valence X Activity Type interaction separately for each level of Target Type (Other vs. Self). For Other participants, the three-way interaction was not significant, *F*(1, 72) < 1, *p* < .99; for Self participants, it was, *F*(1, 72) = 6.13, *p* < .02. To further clarify our findings, we then decomposed this significant three-way interaction for Self participants in two ways.

First, we examined the two-way Cognitive Activity X Trait Valence interaction for each level of Activity Type (Mental vs. Written). For Mental participants, the interaction was not significant, *F*(1, 36) < 1; for Written participants, it was, *F*(1, 36) = 8.12, *p* < .02. In terms of simple effects, Explanatory participants (who wrote down their inquiries) endorsed positive traits significantly less strongly than Descriptive participants (who wrote down their thoughts), *F*(1, 18) = 6.92, *p* < .02; they also endorsed negative traits marginally more strongly, *F*(1, 18) = 2.14, *p* < .14. As predicted, self-enhancement curtailment occurred only when self-directed explanatory inquiries took written form.

Second, we examined the two-way Activity Type X Trait Valence interaction for each level of Cognitive Activity (Descriptive vs. Explanatory). For Descriptive participants, the interaction was not significant, *F*(1, 36) < 1; for Explanatory participants, it was, *F*(1, 36) = 10.92, *p* < .002. In terms of simple effects, Written participants (who wrote down why they did or did not possess traits) endorsed positive traits less strongly than Mental participants (who merely contemplated why they did or did not possess traits), *F*(1, 18) = 4.78 , *p* < .05; they also endorsed negative traits more strongly, *F*(1, 18) = 6.14, *p* < .05. As predicted, self-enhancement curtailment occurred only when self-directed writings documented reasons for possessing or lacking traits.

In summary, we confirmed all hypotheses regarding the boundary conditions of the effects observed in Experiment 1. Self-enhancement was curtailed when participants (a) considered their own traits rather than those of another person, (b) wrote down what they considered rather than merely keeping it in mind, and (c) inquired into why those traits were held as opposed to the extent to which they were held.

Supplementary analyses. With a view to replicating the results of Experiment 1 and more robustly testing our hypotheses, we conducted additional planned comparisons between experimental and control participants. In particular, we examined three types of participants: (a) those who explanatorily introspected about their own personality traits in written form (Self/Written/Explanatory, or SWC); (b) those who reflected upon the extent of their own personality traits in written form (Self/Written/Descriptive, or SWD); and (c) those who considered possible uses for three everyday objects in written form (Control, or CON). We principally sought to investigate whether SWC participants self-enhanced less than CON participants, replicating the results of Experiment 1. However, we additionally sought to investigate whether (a) the SWD and CON participants self-enhanced similarly with one another, but (b) together self-enhanced more than SWC participants. This would establish the essential comparability of the Descriptive Introspection manipulation (newly featured in Experiment 2) and the Control manipulation (also featured in Experiment 1). Any effects of explanatory introspection would therefore be tested relative to a consistent baseline in Experiments 1 and 2.

We duly regressed the composite self-descriptiveness index onto three predictors: a main effect contrast for Trait Valence (Positive = 1, Negative = -1); two main effects contrasts to test predictions (a) and (b) above respectively [(a) SWC = 0, SWD = +1, CON = -1; (b) SWC = 1, SWD = -.5, CON = -.5]; and two interaction contrasts created by multiplying the contrast values for the Trait Valence main effect by the contrast values for each of the Cognitive Activity main effects. All relevant means and standard deviations are presented in Table 3.

First, we compared SWC participants to SWD and CON participants combined in terms of their Trait Valence differentials. The critical interaction contrast was significant, $B = -.21$, $t(54) = -3.67$, $p < .001$. Next, we conducted both main effect contrasts for Positive and Negative traits separately. As predicted, the difference between SWD and CON participants was not significant for Positive traits, $B = .06$, $t(27) = .37$, $p < .75$, or for Negative traits, $B = .08$, $t(27) = .43$, $p < .65$. These results attest to the comparability of the Descriptive and Control introspection conditions. Also as predicted, the difference between SWC participants, and the SWD and CON participants combined, was significant for both Positive traits, $B = -.51$, $t(27) = -3.11$, $p < .01$, and Negative Traits, $B = .39$, $t(27) = 2.21$, $p < .05$. Self-enhancement was significantly curtailed among SWC participants relative to SWD and CON participants.

Reasons

We will start by providing examples of reasons that participants listed in the Cognitive Activity (Explanatory vs. Descriptive) X 2 (Trait Valence: Positive vs. Negative) conditions, when the target type was the self and the activity type was written. These examples are: “I am always straightforward and tell a person how it is” (confirming honest, Explanatory Positive condition); “Sometimes I tell people things that others don’t want me to tell them” (confirming untrustworthy, Explanatory Negative condition); “People always tell me how nice I am” (confirming kind, Descriptive Positive condition); and “It is too tiring to be nice all the time” (confirming unkind, Descriptive Negative condition).

In Experiment 1, Explanatory participants rated (following the manipulation) the degree to which each trait was self-descriptive, and then labeled the reasons they had listed as either confirming or disconfirming each trait. However, this practice was vulnerable to confounds involving self-perception (Bem, 1972) or dissonance (Festinger & Carlsmith, 1959). That is, participants’ reasons-labeling decisions may have been driven, at least in part, by a need to maintain consistency with the prior self-descriptiveness ratings. For example, participants who rated themselves as honest may subsequently have come to perceive the reasons they listed as confirming their honesty, especially if they valued being honest, or their reasons admitted of interpretation.

To partly address this possibility, we asked two independent coders, unaware of the hypotheses under study, to label each reason that Explanatory participants listed as either confirming or disconfirming each relevant trait (for either Self or Other). The coders agreed 96% of the time and resolved disagreements though discussion. We proceeded by computing a confirmation index for each participant ($\alpha = .81$) as in Experiment 1. Next, we entered this index into a Target Type X Trait Valence ANOVA. Replicating Experiment 1, participants were more likely to generate reasons confirming positive traits than reasons confirming negative traits (*M*s = .86 vs. .27), $F(1, 25) = 36.90$, $p < .001$.[3] However, this effect was qualified by an interaction, $F(1, 25) = 5.75$, $p < .02$. Participants showed a weak explanatory tendency to confirm positive traits less for Self ($M = .80$) than for Other ($M = .92$), $F(1, 12) = 1.64$, $p < .22$, combined with a marginal tendency confirm negative traits more for Self ($M = .46$) than for Other ($M = .15$), $F(1, 13) = 4.24$, $p < .06$. In our view, this makes it less likely that consistency motivation led participants to revise their reason-labels in light of their self-descriptiveness ratings. If they had, then the tendency to confirm positive and disconfirm negative traits should have been more pronounced in the more personally consequential Self condition than in the less personally consequential Other condition.

As in Experiment 1, we examined the relation between participants' self-descriptiveness ratings and the confirmation index derived from participants' own reason-labelings. The correlation was again significant, $r(27) = .85$, p < .001, suggesting that participants partially based their self-descriptiveness ratings on the reasons that they generated, although the reverse causal path cannot be ruled out. As before, no significant difference emerged in participants' propensity to form online self-evaluations ($z = .10$, $p < .92$) after explanatorily introspecting about positive traits, $r(12) = .63$, $p < .02$, and after explanatorily introspecting about negative traits, $r(13) = .61$, $p < .02$.

Summary

Experiment 2 achieved several substantive objectives. First, it replicated the self-enhancement curtailment effect observed in Experiment 1. Second, it ruled out a potential rival explanation for the effect, namely, that it was merely due to more protracted thinking. Third, Experiment 2 extended Experiment 1 by identifying several key boundary conditions of the self-enhancement curtailment effect. It showed that explanatory cognition is essential (descriptive cognition does not suffice); it showed that self-directed cognition is essential (other-directed cognition does not suffice); and it showed that that written expression is essential (abstract contemplation does not suffice). Finally, Experiment 2 provided further correlations between listed-reasons and self-ratings suggesting that the changes in the acute accessibility of self-knowledge lie at the heart of the self-enhancement curtailment effect.

Experiment 3

In Experiment 3, we sought to test whether explanatory introspection curtails self-enhancement by reducing self-certainty (Petty et al., 2002). The experiment followed a five-step procedure. First, participants rated themselves on three positive traits. (For simplicity, we omitted negative traits). We labeled these ratings pre-introspection self-descriptiveness, or SD_{PRE}. Second, we introduced the manipulation: participants were randomly assigned to introspect explanatorily, to introspect descriptively, or to perform a control task. Third, participants rated how certain they were that they possessed the three positive traits; that is, they indicated how sure they were about SD_{PRE}. We labeled these ratings pre-introspection self-description certainty, or $CERT_{PRE}$. Fourth, participants re-rated themselves on the same three traits. We labeled these ratings post-introspection self-descriptiveness, or SD_{POST}. (This dependent measure corresponds to the main dependent measure of Experiments 1 and 2.) Fifth, participants re-rated how certain they were that they possessed the three positive traits; that is, they indicated afresh how sure they were about SD_{POST}. We labeled these ratings post-introspection self-description certainty, or $CERT_{POST}$.

What pattern of results would suggest that a reduction in self-certainty was responsible for the impact of explanatory introspection on self-enhancement? Just this: After explanatorily introspecting, participants should be relatively less certain about their original self-views. This decrease in certainty should in turn shape their post-manipulation self-views, now revised downwards. However, after re-expressing their revised self-views, participants' self-certainty should rebound.

In more technical terms, we expected that Explanatory participants (relative to both Descriptive and Control participants) would, following the manipulation, have lower CERT_{PRE} ratings, because they would now be less certain of their original self-views. Such participants would also have lower SD_{POST} ratings, controlling for SD_{PRE} ratings, because explanatory introspection would have curtailed their proclivity to self-enhance. Most importantly, variations in self-certainty would also mediate the effects of the manipulation on self-views; that is, CERT_{PRE} ratings would mediate the effects of the manipulation on SD_{POST}. However, following the expression of SD_{POST}, self-certainty would be restored: no differences between conditions in $\text{CERT}_{\text{POST}}$ would be observed.

Method

Participants, Experimental Design, and Procedure

Fifty-one participants were assigned randomly to one of three conditions: explanatory introspection (Explanatory), descriptive introspection (Descriptive), and object-use generation (Control). Thus, the experiment featured a one-way balanced between-subjects design. Procedures were largely identical to those of Experiment 2 (in the Self and Written conditions). As in Experiment 1, all participants left the entire booklet behind.

Participants completed SD_{PRE} ratings for three traits: honest, kind, and trustworthy. The manipulation followed. Finally, all participants completed CERT_{PRE} ratings, SD_{POST} ratings, and $\text{CERT}_{\text{POST}}$ ratings.

Measures

SD_{PRE} ratings. Participants responded to two items for each trait. The first read "Please rate yourself, relative to *other college students your own age*, on the trait ___" (1 = *lower 5%*, 10 = *upper 5%*). The second read, "Please rate yourself, relative *to other people in general*, on the trait ___" (1 = *lower 5%*, 10 = *upper 5%*). We averaged both items for each trait to create three indices, (α = .91, .85, and .95, for honest, kind, and trustworthy, respectively). Next, we averaged these indices to create a final composite index, SD_{PRE} (α = .85). Higher scores indicate higher pre-manipulation levels of trait self-descriptiveness.

$CERT_{PRE}$. Participants responded to three items for each trait. The first read, "How *certain* are you of the accuracy of the ratings you made a few moments ago in reference to the trait ___?" (1 = *not at all certain*, 15 = *very certain*). The second read, "How *confident* are you in the accuracy of the ratings you made a few moments ago in reference to the trait ___?" (1 = *not at all confident*, 15 = *very confident*). The third read, "How *sure* are you that the ratings you made a few moments ago about the trait ___ reflect your true level of the trait ___?" (1 = *not at all sure*, 15 = *very sure*). We averaged the three items for each trait to create three indices (α = .94, .96, and .95, for honest, kind, and trustworthy, respectively). Next, we averaged these indices to create a final composite index, $CERT_{PRE}$ (α = .80). Higher scores indicate greater certainty about pre-manipulation levels of trait self-descriptiveness.

SD_{POST}. Participants responded to three items for each trait. The wording was varied slightly in order to discourage reflexive repetition of previous responses. The first item read, "How descriptive of you is the trait ___?" (1 = *not at all descriptive*, 15 = *very descriptive*). The second read, "To what extent do you think you have the trait ___?" (1 = *not at all*, 15 = *very much*). The third read, "How well does the trait ___ describe you?" (1 = *not well at all*, 15 = *very well*). We averaged the three items for each trait to create three indices (α = .93, .89, and .97, for honest, kind, and trustworthy, respectively), and then averaged these indices to create a final

composite index, *SD_{POST}* ($\alpha = .77$). Higher scores indicate higher levels of post-manipulation trait self-descriptiveness.

$CERT_{POST}$. These items were identical to those used for *$CERT_{PRE}$*, with one minor modification. Each item referred to certainty about the accuracy of "…the ratings you *JUST* made in reference to the trait ___." We averaged the three items for each trait to create three indices ($\alpha = .97$, .98, and .97, for honest, kind, and trustworthy, respectively). Next, we averaged these indices to create a final composite index, *$CERT_{POST}$* ($\alpha = .83$). Higher scores indicate greater certainty about post-manipulation levels of trait self-descriptiveness.

Results and Discussion

Self-Evaluation

All means and standard deviations for the self-evaluation results are presented in Table 4.

Did explanatory introspection reduce self-description certainty? We subjected *$CERT_{PRE}$* ratings to a one-way ANOVA. The main effect was significant, $F(2, 48) = 3.14$, $p < .05$: the pattern suggested that Explanatory participants ($M = 12.03$) were less certain about their traits than both Descriptive participants ($M = 13.03$) and Control ($M = 13.49$) participants. We used planned comparisons to pin down the locus of the effect. Specifically, after standardizing certainty ratings, we devised linear contrasts that (a) compared Explanatory participants to Descriptive and Control participants combined, and (b) compared Descriptive participants to Control participants. We simultaneously entered these orthogonal contrasts as predictors of the standardized certainty ratings. As predicted, Explanatory participants were less self-certain than Descriptive and Control participants combined, $B = -.32$, $t(48) = -2.39$, $p < .03$, but Descriptive and Control participants did not differ in their self-certainty, $B = -.10$, $t(48) = -.77$, $p < .45$. Thus, explanatory introspection reduced certainty about pre-introspection self-descriptiveness ratings.

Did explanatory introspection curtail self-enhancement (after controlling for pre-introspection self-descriptiveness)? We subjected SD_{POST} ratings to a one-way ANCOVA, with SD_{PRE} ratings serving as a covariate. The main effect for the manipulation was again significant, $F(2, 47) = 3.83, p < .05$: the pattern suggested that Explanatory participants ($M = 12.31$) regarded the positive traits as less self-descriptive than both Descriptive participants ($M = 13.14$) and Control participants ($M = 13.62$). Unsurprisingly, the effect of SD_{PRE} ratings on SD_{POST} ratings was also significant, $F(1, 47) = 7.70, p < .01$.

Next, we devised linear contrasts analogous to (a) and (b) described above. We simultaneously entered both contrasts, together with SD_{PRE} ratings, as predictors of SD_{POST}, after again standardizing both sets of ratings. Descriptive and Control participants did not differ in terms of their SD_{POST} ratings, $B = -.16$, $t(47) = -1.29, p < .25$. However, Explanatory participants regarded the positive traits as less self-descriptive than did Descriptive and Control participants combined, $B = -.31$, $t(47) = -2.44, p < .02$. Thus, even after controlling for SD_{PRE} ratings, explanatory introspection curtailed self-enhancement, replicating both previous experiments.

Did self-description certainty statistically mediate the impact of explanatory introspection on self-descriptiveness? To determine whether $CERT_{PRE}$ mediated the impact of explanatory introspection (characterized in terms of the two linear contrasts—[a] and [b] above) on SD_{POST}, we adopted Baron and Kenny's (1986) analytic strategy. We had already satisfied one requirement—that the independent variable should significantly predict the dependent variable. Specifically, we had found that explanatory introspection led to relatively lower SD_{POST} ratings (adjusted for SD_{PRE} ratings). We had also already satisfied another requirement—that the independent variable should significantly predict the proposed mediator. Specifically, we had found that explanatory introspection led to relatively lower $CERT_{PRE}$ ratings. We now sought to satisfy the final requirements—(a) that the proposed mediator, $CERT_{PRE}$ ratings, significantly predict the dependent variable, adjusted *SD_{POST}* ratings, controlling for the independent variable, explanatory introspection and (b) that, in the same analysis, the predictiveness of the independent variable is reduced significantly. We succeeded. Specifically, when adjusted *SD_{POST}* ratings were regressed on *$CERT_{PRE}$* ratings, and on the two linear contrasts (a) and (b), the effect of *$CERT_{PRE}$* ratings persisted, $B = .68$, $t(46) = 7.39$, $p < .001$, but the key linear contrast (a), previously significant, became nonsignificant, $B = -.11$, $t(46) = -1.23$, $p < .25$. Importantly, a significant indirect effect of that contrast on *SD_{POST}* ratings via *$CERT_{PRE}$* emerged, $z = 2.05$, $p < .05$. In summary, the impact of explanatory introspection on post-introspection self-descriptiveness ratings was mediated by certainty about pre-introspection self-descriptiveness ratings.

Was certainty restored following post-introspection self-descriptiveness ratings? We subjected *$CERT_{POST}$* ratings to a one-way ANOVA. Contrary to what was found for *$CERT_{PRE}$* ratings, this main effect was not significant, $F(2, 48) = 1.05$, $p < .40$. Explanatory participants ($M = 12.91$) were nearly as certain about their post-introspection self-descriptiveness ratings as were Descriptive participants ($M = 13.37$) and Control participants ($M = 13.70$). For completeness, we ran the same planned contrasts as before, (a) and (b). Unsurprisingly, neither attained significance: (a) $B = -.19$, $t(48) = -1.31$, $p < .20$; (b) $B = -.09$, $t(48) = -.61$, $p < .60$.

Reasons

Dovetailing the results of Experiments 1 and 2 (for positive traits), explanatory participants generated reasons that they labeled as confirming their self-descriptiveness ratings (77%). However, the design of Experiment 3, unlike that of previous experiments, permitted the disambiguation of two competing causal alternatives: Did Explanatory participants use reasons as a basis for (generating) their self-descriptions? Or did they use their self-descriptions as a basis for (labeling) their reasons? Support for the first alternative would be signaled by (a) a significant positive correlation between the confirmation index and *SD_{POST}* ratings, and (b) no significant positive correlation between the confirmation index and SD_{PRE} ratings. Support for the second alternative would be signaled by the reverse pattern.

Like before, we computed a confirmation index ($\alpha = .71$) and correlated it with SD_{POST} ratings. The correlation was significant, $r(15) = .77, p < .001$. However, the corresponding correlation with SD_{PRE} ratings was not, $r(15) = .37, p < .14$. Moreover, the difference between the two correlations was marginal, $z = 1.66, p < .10$. Thus, a pattern emerged consistent with the first alternative (and with our favored interpretation of relevant findings of Experiments 1 and 2). Explanatory participants based their self-descriptiveness ratings on the products of their introspections, and did not label their reasons in light of their newly revised self-views.

Summary

Experiment 3 established that explanatory introspection curtails self-enhancement by decreasing self-certainty. Three lines of evidence supported this assertion. First, explanatory introspection decreased participants' certainty about their self-views. Second, this decrease in self-certainty fully mediated self-enhancement curtailment. Third, after re-expressing self-views, participants recovered their former levels of self-certainty.

General Discussion

We investigated introspection as a means of curtailing people's natural tendency towards self-enhancement. We began by differentiating between two types of introspection: explanatory and descriptive. People engage in descriptive introspection when they contemplate or describe the extent to which they do or do not possess particular traits: in effect, they consider what kind of person they are. In contrast, people engage in explanatory introspection when they contemplate why they might or might not be a particular kind of person; in effect, they consider the reasons why they are the kind of person they are.

Next, taking our cue from prior research on reasons-analysis (Wilson et al., 1989) and debiasing (Lord et al., 1984), we wondered whether explanatory introspection, as opposed to its descriptive cousin, would curtail self-enhancement. Assuming it did so, we also wondered what the underlying mechanisms might be. We postulated that participants who explanatorily introspect conduct an autobiographical memory search for behavioral instances that support or refute the possession of trait (i.e., “reasons”). Retrieved instances then alter the accessibility of some items of self-knowledge. Because self-views are based in part on accessible self-knowledge (Fazio et al., 1981), they consequently undergo at least temporary modification (cf. Wilson et al., 1989). Moreover, given that introspected traits are themselves either positive (e.g., kind) or negative (e.g., selfish), some reasons generated will be relatively congenial (supporting positive traits or refuting negative ones), whereas others will be relatively uncongenial (supporting negative traits or refuting positive ones). Although the former should prevail numerically—yet another example of self-enhancement—the latter should nonetheless carry more weight (cf. Baumeister et al., 2001). Hence, self-views should become more moderate, with positive traits being endorsed less strongly, and negative traits more strongly. In addition, explanatory introspection should leave an experiential mark: a heightened state of uncertainty about self-views. Indeed, we postulated that this increase in self-uncertainty would mediate the moderating effects of explanatory introspection on self-enhancement.

To test our hypotheses, we conducted three experiments. In all three, participants considered a set of central traits in one way or another, and then indicated the extent to which those traits characterized them. Experiment 1 established that explanatory introspection curtails self-enhancement. Participants who asked themselves why they did or did not possess traits were less likely to endorse positive traits, and (marginally) more likely to endorse negative traits. Moreover, participants’ deflated self-evaluations covaried with the confirmatory reasons they generated, implicating a role for alterations in self-knowledge accessibility. Experiment 2 replicated, clarified, and extended the findings of Experiment 1. For self-enhancement to be curtailed, participants' trait-related inquiries had to be explanatory (not descriptive), self-directed (not other-directed), and transcribed (not just contemplated). Finally, Experiment 3 provided evidence that reductions in self-certainty mediate the impact of explanatory introspection on self-enhancement. It also provided evidence that participants more probably based their self-descriptions on the reasons that they generated than retrospectively classified the reasons they generated in light of their self-descriptions.

One general observation is worth making with respect to our findings. First, although explanatory introspection curtailed self-enhancement significantly, the magnitude of its impact was modest. In particular, explanatory introspection participants still rated positive traits as more self-descriptive than negative traits in an absolute sense; for example, on a 15-point scale, the respective *M*s were 12.20 vs. 3.68 (Experiment 1) and 10.80 vs. 5.79 (Experiment 2, Written Activity Type). Yet this is hardly surprising, for two reasons. First, the propensity to self-enhance, being so ingrained, is difficult to dislodge completely (Sedikides & Gregg, 2003). Second, the fact that people possess a rich fund of knowledge about self (Higgins, 1996) is liable to make self-views relatively resistant to explanatory inquiry. It has been found, for instance, both in classic research on reasons analysis (Wilson, Kraft, & Dunn, 1989), as well as in more recent research on value change (Bernard, Maio, & Olson, 2003b), that the perturbing effects of explanatory introspection fade when people's attitudes or values, the intended targets of change, are cognitively well-supported. Moreover, central traits, being valued parts of one's identity, are liable to be particularly well cognitively supported (Markus, 1977; Sedikides, 1995). Nevertheless, we consistently found that explanatory introspection curtailed self-enhancement even when central traits were pondered. Perhaps the self, being an object of special interest, elicits particularly elaborate cognitive processing, sufficient to modify its more elaborate structure (Greenwald & Banaji, 1989). However, we surmise that the impact of explanatory introspection might be yet more pronounced when peripheral traits are pondered, subject to the caveat that self-views on peripheral traits will initially be less extreme (Sedikides, 1993, 1995).

We would also like to address a potential limitation of our research that pertains to a boundary condition in Experiment 2. In particular, participants in the Mental condition of Activity Type were instructed to take a few minutes to think about reasons. In this control condition, the effects of explanatory introspection were absent, compared to the experimental (Written) condition, where they were present. Although informal observation and exit interviews satisfied us that participants in the Mental condition took the task seriously (i.e., they seemed attentive to instructions and contemplative during the allotted introspection time), we are unable to back up our claim with a manipulation check. Nevertheless, we wish to point out that, in Experiment 2, we did show that explanatory introspection (i.e., writing reasons why one does or does not possess various traits) curtailed self-enhancement relative to descriptive introspection (i.e., describing the extent to which one does or does not possess various traits)—and that this was, theoretically speaking, the most critical finding. Moreover, this finding was conceptually replicated: in Experiment 1, explanatory introspection curtailed self-enhancement relative to a control condition, and, in Experiment 3, explanatory introspection curtailed self-enhancement relative to descriptive introspection. The validity of Experiment 2 results is further bolstered by the finding that the highest reduction in self-enhancement was observed when participants (a) introspected explanatorily, (b) about the self, and (c) listed reasons.

Raising and Lowering Self-Esteem

Empirical documentations of self-enhancement abound. Individuals both affirm (Kumashiro & Sedikides, 2005; Steele, 1988) and protect (Sedikides, Green, & Pinter; 2004;

Tesser, 2001) their valued self-views with fervor and ingenuity. Happily, self-enhancement affords many intrapsychic benefits (Taylor & Brown, 1988). Sadly, it is also carries several costs, both intrapsychic and interpersonal (Robins & Beer, 2001). It follows that keeping self-enhancement in check, although it may entail some intrapsychic drawbacks, may also furnish some intrapsychic and interpersonal advantages.

Traditionally, much effort has been expended to raise self-esteem—that is, making self-enhancement the dispositional default (Sedikides & Gregg, 2003). Although self-help gurus have spearheaded this effort by penning self-help books for mass consumption (Branden, 1995; McKay & Fanning, 2000), academic psychologists have made contributions of their own, most recently pioneering subtle associative techniques (Baccus, Baldwin, & Packer, 2004; Dijksterhuis, 2004). The drive to raise self-esteem, whether successful or not, has been premised on the assumption that high self-esteem is a decidedly desirable psychological characteristic that has primarily prosocial implications (California Task Force to Promote Self-Esteem and Personal and Social Responsibility, 1990). However, this assumption is suspect (Baumeister, Campbell, Krueger, & Vohs, 2003). Although high self-esteem may feel good subjectively, it does not appear to be a prescription for objective achievement or social harmony (although see Donnellan et al., 2005). Indeed, there are several reasons why not having a maximally positive self-view might be advantageous (Sedikides, Gregg, & Hart, in press). First, compared to blatant self-enhancers, people with moderate and balanced self-views are better liked, both as individuals (Robinson, Johnson, & Shields, 1995) and as work colleagues (Wosinska, Dabul, Whetstone-Dion, & Cialdini, 1996). In addition, people with particularly inflated self-views (e.g., narcissists) are interpersonally abrasive rather than constructive (Sedikides, Rudich, Gregg, Kumashiro, & Rusbult, 2004). Finally, a general but powerful argument against self-enhancement is that it hampers accurate self-assessment (Duval & Silvia, 2002; Wilson & Dunn, 2004), leading to overconfidence that impairs the quality of decision-making in such consequential domains as health, education, and business (Dunning, Heath, & Suls, 2004).

We do not wish to argue that self-effacement is better than self-enhancement, or that all attempts to raise self-esteem are fundamentally wrongheaded. Rather, we wish to argue that both self-effacement and self-enhancement have distinctive advantages and disadvantages—perhaps inextricably intertwined (Sedikides & Luke, in press). This being the case, raising self-esteem will be more desirable in some contexts, and lowering self-esteem in others: it all depends on whether the advantages outweigh the disadvantages.

Of course, explanatory introspection can occur not only in response to instruction, but also in everyday life spontaneously. We consider below two possible contexts in which explanatory introspection might play a role, with concurrent effects on self-certainty. In one case, explanatory introspection takes the form of a deliberate intervention intended to be beneficial. In another case, it takes the form of naturally occurring phenomenon liable to cause harm.

Explanatory introspection as a tonic for narcissism. By definition, narcissists[4] self-aggrandize, that is, engage in excessive self-enhancement. For example, they deny possessing commonplace flaws (Paulhus, 1998), objectively overestimate their intelligence (Farwell & Wohlwend-Lloyd, 1998), and regard themselves as more influential and attractive than others do (John & Robins, 1994). Such illusions, being pronounced, put them at special risk of error when it comes to making important decisions (Dunning et al., 2004). In tandem, narcissists cause trouble for others, perhaps as a direct result of their inflated but somewhat fragile egos (Sedikides et al., 2004). For example, they put down those who outdo them (Kernis & Sun, 1994; Morf & Rhodewalt, 1993), punish those who criticize them (Bushman & Baumeister, 1998), and treat their intimate partners casually (Campbell, Foster, & Finkel, 2002). It follows that reducing their self-esteem might have salutary effects, both intrapersonally, by fostering cognitive realism, and interpersonally, by fostering harmonious relationships.

Unfortunately, narcissists doggedly self-regulate to avoid the possibility of self-effacement (Morf & Rhodewalt, 2001). Hence, the strategy of explicitly confronting them with shortcomings is liable not only not to work, but also to backfire. A more unobtrusive approach is therefore called for. In this connection, invitations to introspect explanatorily may fit the bill. For example, narcissists might be prepared to consider in writing the reasons why they do or do not possess a particular set of traits, permitting a dent to be made in their robust levels of self-certainty (Rhodewalt & Regalado, 2000). Of course, it is unrealistic to expect that such an approach would have a long-lasting impact on narcissists, especially given the small effects obtained in our research. At best, the extent and durability of any changes would be an empirical question and would depend upon the precise methodology used.

Explanatory introspection as a preserver of low self-esteem. Researchers have puzzled over the persistence of low self-esteem. Why does it not reliably recede when there is objective reason to feel proud or positive feedback from others? Several hypotheses have been put forward, and some have received empirical support. For instance, people with low self-esteem do not find their own self-generated positive feedback credible (Josephs, Bosson, & Jacobs, 2003). They also lack the energy to engage in mood repair activities, even when they expect them to work (Heimpel, Wood, Marshall, & Brown, 2002). It has even been suggested that people with low self-esteem do not desire positive feedback because of the threat it poses to the coherence of their identity (Swann, Rentrow, & Guinn, 2003).

We suggest that yet another factor is involved: habitual explanatory introspection. We propose that people with low self-esteem keep attempting to explain why they are the way they are because the way they are dissatisfies them.[5] Hence, they continually undermine their capacity to self-enhance. Although we could not locate any direct evidence for this contention, there are several lines of indirect evidence consistent with it. First, it is already known that other varieties of cognitive activity, such as counterfactual reasoning, vary with levels of self-esteem (Roese & Olson, 1993). Second, the self-conceptions of people with low self-esteem are known to be more tentative and less coherent (Campbell, 1990). This is precisely what one would expect if self-certainty was being reduced via repeated explanatory introspection. Third, explanatory attribution for events related to the self is greater when those events are negative (Weiner, 1985). Given that people with low self-esteem appraise themselves and their attributes negatively (Baumeister et al., 2003) and experience higher levels of negative affect (Leary & McDonald, 2003), it would hardly be surprising if they also sought explanations for these negative "events." Admittedly, such enquiries would be conducted without the aid of pen and paper, a precondition for curtailing self-enhancement according to Experiment 2. However, it may simply be a matter of dosage: if people with low self-esteem explanatorily introspect in their own minds with sufficient frequency and intensity, and if they seek reasons for the same problematic traits over and over again, then no pen and paper may be needed to bring about the required alterations in the accessibility of self-knowledge. People high in private self-consciousness (Fenigstein, Scheier, & Buss, 1975) and in self-doubt (Oleson, Poehlmann, Yost, Lynch, & Arkin, 2000) may be similarly susceptible to spontaneous explanatory introspection and suffer the consequences.

Coda

Asking oneself why one might or might not possess particular traits moderates self-evaluations by reducing certainty about these traits. This finding suggests a new take on Socrates' famous dictum that "the unexamined life is not worth living" (Loomis, 1942, p. 56). If asking this "why" question of oneself lowers self-enhancement, then the results are liable to be subjectively unpleasant. Moreover, if one's propensity to self-enhance is already chronically low, then the results may also be objectively counterproductive. If so, then the examined life would be less worth living, not more. On the other hand, if one’s propensity to self-enhance is excessive, then a dose of explanatory introspection may be just what the doctor ordered. Subjectively, it may not make one’s own life any more worth living. However, by curtailing one’s own egotism, it may improve the lives of those with whom one interacts.

References

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

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

Anderson, C. A. (1982). Inoculation and counter-explanation: Debiasing techniques in the perseverance of social theories. *Social Cognition, 1*, 126-139.

Anderson, C. A., Lepper, M. R., & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology, 39*, 1037-1049.

Baccus, J. R., Baldwin, M. W., & Packer, D. J. (2004). Increasing implicit self-esteem through classical conditioning. *Psychological Science*, *15*, 498-502.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.

Baumeister, R. F. (1998). The self. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 1, pp. 680-740). New York: McGraw-Hill.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5*, 323-370.

Baumeister, R. F., Campbell, J. D., Krueger, J. I., & Vohs, K. D. (2003). Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles? *Psychological Science in the Public Interest, 4*, 1-44.

Baumeister, R. F., Heatherton, T. F., & Tice, D. M. (1993). When ego threats lead to self-regulation failure: Negative consequences of high self-esteem. *Journal of Personality and Social Psychology, 64*, 141-156.

Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6, pp. 1-62). New York: Academic Press.

Bernard, M. K., Maio, G. R., & Olson, J. M. (2003a). Effects of introspection about reasons for values: Extending research on values-as-truisms. *Social Cognition, 21*, 1-25.

Bernard, M. K., Maio, G. R., & Olson, J. M. (2003b). The vulnerability of values to attack: Inoculation of values and value-relevant attitudes. *Personality and Social Psychology Bulletin, 29*, 63-75.

Bless, H., & Forgas, J. P. (2000). *The message within: The role of subjective experience in social cognition and behavior*. Philadelphia, PA: Psychology Press.

Bonanno, G. A., Rennicke, C., & Dekel, S. (2005). Self-enhancement among high-exposure survivors of the September 11th terrorist attack: Resilience or social maladjustment? *Journal of Personality and Social Psychology, 88*, 984-998.

Branden, N. (1995). *Six pillars of self-esteem*. New York: Bantam.

Bushman, B. J., & Baumeister, R. F. (1998). Threatened egotism, narcissism, self-esteem, and direct and displaced aggression: Does self-love or self-hate lead to violence? *Journal of Personality and Social Psychology*, *75*, 219-229.

California Task Force to Promote Self-Esteem and Personal and Social Responsibility (1990). *Toward a state of self-esteem*. Sacramento, CA: California State Department of Education.

Campbell, J. D. (1990). Self-esteem and clarity of the self-concept. *Journal of Personality and Social Psychology*, *59*, 538-549.

Campbell, W. K., Foster, C. A., & Finkel, E. J. (2002). Does self-love lead to love for others?: A story of narcissistic game-playing. *Journal of Personality and Social Psychology*, *83*, 340-354.

Caroll, J. S. (1978). The effect of imagining an event on expectations for the event: An interpretation in terms of the availability heuristic. *Journal of Experimental Social Psychology, 14*, 88-96.

Davies, M. F. (2003). Confirmatory bias in the evaluation of personality descriptions: Positive test strategies and output inferences. *Journal of Personality and Social Psychology, 85*, 736-744.

Dauenheimer, D. G., Stahlberg, D., Spreeman, S., & Sedikides, C. (2002). Self-enhancement, self-assessment, or self-verification?: The intricate role of trait modifiability in the self-evaluation process. *Revue Internationale De Psychologie Sociale, 15*, 89-112.

Dijksterhuis, A. (2004). I like myself but I don't know why: Enhancing implicit self-esteem by subliminal evaluative conditioning. *Journal of Personality and Social Psychology*, *86*, 345-355.

Donnellan, M. B., Trzesniewski, K. H., Robins, R. W., Moffitt, T. E., & Caspi, A. (2005). Low self-esteem is linked to antisocial behavior and delinquency. *Psychological Science*, 16, 328-335.

Dunning, D. A. (1999). A newer look: Motivated social cognition and the schematic representation of social concepts. *Psychological Inquiry, 10*, 1-11.

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.

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment. *Psychological science in the public interest*, *5*, 69-106.

Duval, T. S., & Silvia, P. J. (2002). Self-awareness, probability of improvement, and the self-serving bias. *Journal of Personality and Social Psychology, 82*, 49-61.

Epley, N., & Dunning, D. (2000). Feeling “holier than thou”: Are self-serving assessments produced by errors in self- or social prediction? *Journal of Personality and Social Psychology, 79*, 861-875.

Farwell, L., & Wohlwend-Lloyd, R. (1998). Narcissistic processes: Optimistic expectations, favorable self-evaluations, and self-enhancing attributions. *Journal of Personality*, *66*, 65-83.

Fazio, R. H., Effrein, E. A., & Falender, V. J. (1981). Self-perceptions following social interaction. *Journal of Personality and Social Psychology, 41*, 232-242.

Fenigstein, A., Scheier, M. F., & Buss, A. H. (1975). Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology, 43*, 522-527.

Festinger, A. & Carlsmith, J. (1959). Cognitive consequences of forced compliance. *Journal of*

Personality and Social Psychology, *58*, 203-10.

Gosling, S. D., John, O. P., Craik, K. H., & Robins, R. W. (1998). Do people know how they behave? Self-reported act frequencies compared with on-line coding by observers. *Journal of Personality and Social Psychology, 74*, 1337-1349.

Cottingham, J., Stoothoff, R., & Murdoch, D. (Eds.). (1984). *The philosophical writings of Descartes* (Vols 1-3). Cambridge, UK: Cambridge University Press.

Greenwald, A. G. (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist*, *35*, 603-618.

Greenwald, A. G., and Banaji, M. R. (1989). The self as a memory system: Powerful, but ordinary. *Journal of Personality and Social Psychology*, *57*, 41-54.

Heimpel, S. A., Wood, J. V., Marshall, M. A., & Brown, J. D. (2002). Do people with low self-esteem really want to feel better? Self-esteem differences in motivation to repair negative moods. *Journal of Personality and Social Psychology*, *82*, 128-147.

Higgins, E. T. (1996). The "self digest": Self-knowledge serving self-regulatory functions. *Journal of Personality and Social Psychology, 71*, 1062-1083.

Hirt, E. R., Kardes, F. R., & Markman, K. D. (2004). Activating a mental simulation mind-set through generation of alternatives: Implications for debiasing in related and unrelated domains. *Journal of Experimental Social Psychology, 40*, 374-383.

Hirt, E. R., & Markman, K. D. (1995). Multiple explanation: A consider-an-alternative strategy for debiasing judgments. *Journal of Personality and Social Psychology, 69*, 1069-1086.

Hixon, J. G., & Swann, W. B. (1993). When does introspection bear fruit? Self-reflection, self-insight, and interpersonal choices. *Journal of Personality and Social Psychology, 64*, 35-43.

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

Josephs, R. A., Bosson, J. K., & Jacobs, C. G. (2003). Self-esteem maintenance processes: Why low self-esteem may be resistant to change. *Personality & Social Psychology Bulletin*, *29*, 920-933.

Kernis, M. H., & Sun, C. (1994). Narcissism and reactions to interpersonal feedback. *Journal of Research in Personality*, *28*, 4-13.

Kihlstrom, J. F., Beer, J. S., & Klein, S. B. (2003). Self and identity as memory. In M. R. Leary & J. P. Tangeny (eds.), *Handbook of self and identity* (pp. 68-90). New York: Gilford.

Klein, S. B., & Loftus, J. (1988). The nature of self-referent encoding: The contributions of elaborative and organizational processes. *Journal of Personality and Social Psychology, 55*, 5-11.

Koehler, D. J. (1991). Explanation, imagination, and confidence in judgment. *Psychological Bulletin, 110*, 499-519.

Kahneman, D., & Tversky, A. (1981). The framing of decisions and the rationality of choice. *Science*, *221*, 453-458.

Kumashiro, M., & Sedikides, C. (2005). Taking on board liability-focused feedback: Close positive relationships as a self-bolstering resource. *Psychological Science, 16*, 732-739.

Kunda, Z., & Sanitioso, R. (1989). Motivated changes in the self-concept. *Journal of Experimental Social Psychology, 25*, 272-285.

Leary, M. R., & MacDonald, G. (2003). Individual differences in self-esteem: A review and theoretical integration. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (pp. 401-418) New York: Guilford Press.

Leary, M. R., Tchividjian, L. R., & Kraxberger, B. E. (1994). Self-presentation can be hazardous to your health: Impression management and health risk. *Health Psychology, 13*, 461-470.

Loomis, L. R. (1942) (Ed.). *Plato: Apology, Crito, Phaedo, Symposium, Republic* (B. Jowett, Trans.). New York: Walter J. Black.

Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology, 47*, 1231-1243.

Maio, G. R., & Olson, J. M. (1998). Values as truisms: Evidence and implications. *Journal of Personality and Social Psychology, 74*, 294-311.

Maio, G. R., Olson, J. M., Allen, L., & Bernard, M. M. (2001). Addressing discrepancies between values and behavior: The motivating effects of reasons. *Journal of Experimental Social Psychology, 37*, 104-117.

Markman, K. D., & Hirt, E. R. (2002). Social prediction and the "allegiance bias." *Social Cognition, 20*, 58-86.

Markus, H. (1977). Self-schemata and processing information about the self. *Journal of Personality and Social Psychology*, *35*, 63-78.

McKay, M., & Fanning, P. (2000). *Self-esteem: A proven program of cognitive techniques for assessing, improving, and maintaining your self-esteem* (3rd Ed.). Oakland, CA: New Harbinger Publications.

Mezulis, A. H., Abramson, L. Y., Hyde, J. S., & Hankin, B. L. (2004). Is there a universal positivity bias in attributions?: A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychological Bulletin, 130*, 711-747.

Morf, C. C., & Rhodewalt, F. (1993). Narcissism and self-evaluation maintenance: Explorations in object relations. *Personality and Social Psychology Bulletin*, *19*, 668-676.

Morf, C. C., & Rhodewalt, F. (2001). Unraveling the paradoxes of narcissism: A dynamic self-regulatory processing model. *Psychological Inquiry*, *12*, 177-196.

Oettingen, G., & Gollwitzer, P. M. (2001). Goal setting and goal striving. In A. Tesser & N. Schwarz (Eds.), *Intraindividual processes: Blackwell Handbook of Psychology* (pp. 329-347). Malden, Mass: Blackwell Publishers, Inc.

Oleson, K. C., Poehlmann, K. M., Yost, J. H., Lynch, M. E., & Arkin, R. M. (2000). Subjective overachievement: Individual differences in self-doubt and concern with performance. *Journal of Personality, 68*, 491-524.

Paulhus, D. L. (1998). Interpersonal and intrapsychic adaptiveness of trait self-enhancement: A mixed blessing? *Journal of Personality and Social Psychology, 74*, 1197-1208.

Pennebaker, J.W. (2003). Writing about emotional experiences as a therapeutic process. In P. Salovey, J.A. Rothman et al. (Eds.), *Social psychology of health: Key readings in social psychology* (pp 362-368). New York, NY: Psychology Press.

Petty, R. E., Brinol, P., & Tormala, Z. L. (2002). Thought confidence as a determinant of persuasion: The self-validation hypothesis. *Journal of Personality and Social Psychology, 82*, 722-741.

Pronin, E., Yin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin, 3*, 369-381.

Raskin, R., & Hall, C. S. (1979). A narcissistic personality inventory. *Psychological Reports*, *45*, 590.

Rhodewalt, F., & Regalado, M. (2000). *NPI Narcissism and the Structure of the Self.* Unpublished data, University of Utah.

Robins, R. W., & Beer, J. S. (2001). Positive illusions about the self: Short-term benefits and long-term costs. *Journal of Personality and Social Psychology, 80*, 340-352.

Robinson, M. D., Johnson, J. T., & Shields, S. A. (1995). On the advantages of modesty: The benefits of a balanced self-presentation. *Communication Research*, *22*, 575-591.

Roese, N. J., & Olson, J. M. (1993). Self-esteem and counterfactual thinking. *Journal of Personality and Social Psychology*, *65*, 199-206.

Ross, L., Lepper, M. R., Strack, F., & Steinmetz, J. (1977). Social explanation and social expectation: Effects of real and hypothetical explanations on subjective likelihood. *Journal of Personality and Social Psychology, 35*, 817-829.

Sanitioso, R., Kunda, Z., & Fong, G. T. (1990). Motivated recruitment of autobiographical memories. *Journal of Personality and Social Psychology, 59*, 229-241.

Sanna, L. J. (2000). Mental simulation, affect, and personality: A conceptual framework. *Current Directions in Psychological Science, 9*, 168-173.

Schlenker, B. R., & Leary, M. R. (1982). Audiences' reactions to self-enhancing, self-denigrating, and accurate self-presentations. *Journal of Experimental Social Psychology, 18*, 89-104.

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simmons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology, 61*, 195-202.

Sedikides, C. (1993). Assessment, enhancement, and verification determinants of the self-evaluation process. *Journal of Personality and Social Psychology, 65*, 317-338.

Sedikides, C. (1995). Central and peripheral self-conceptions are differentially influenced by mood: Tests of the differential sensitivity hypothesis. *Journal of Personality and Social Psychology, 69*, 759-777.

Sedikides, C., Campbell, W. K., Reeder, G., & Elliot, A. J. (1998). The self-serving bias in relational context. *Journal of Personality and Social Psychology, 74*, 378-386.

Sedikides, C., Campbell, W. K., Reeder, G., Elliot, A. J., & Gregg, A. P. (2002). Do others bring out the worst in narcissists? The "Others Exist for Me" illusion. In Y. Kashima, M. Foddy, & M. Platow (Eds.), *Self and identity: Personal, social, and symbolic* (pp. 103-123). Mahwah, NJ: Lawrence Erlbaum Associates.

Sedikides, C., Gaertner, L., & Toguchi, Y. (2003). Pancultural self-enhancement. *Journal of Personality and Social Psychology, 84*, 60-70.

Sedikides, C., & Green, J. D. (2000). On the self-protective nature of inconsistency/negativity management: Using the person memory paradigm to examine self-referent memory. *Journal of Personality and Social Psychology, 79*, 906-922.

Sedikides, C., & Green, J. D. (2004). What I don't recall can't hurt me: Information negativity versus information inconsistency as determinants of memorial self-defense. *Social Cognition, 22*, 4-29.

Sedikides, C., Green, J. D., & Pinter, B. (2004). Self-protective memory. In D. R. Beike, J. M. Lampinen, & D. A. Behrend (Eds.), *The self and memory* (pp. 161-179). Philadelphia, PA: Psychology Press.

Sedikides, C., & Gregg. A. P. (2003). Portraits of the self. In M. A. Hogg & J. Cooper (Eds.), *Sage handbook of social psychology* (pp. 110-138). London: Sage Publications.

Sedikides, C., Gregg, A. P., & Hart, C. M. (in press). The importance of being modest. In C. Sedikides & S. Spencer (Eds.), *Frontiers in social psychology: The self*. Philadelphia, PA: Psychology Press.

Sedikides, C., Herbst, K. C., Hardin, D. P., & Dardis, G. J. (2002). Accountability as a deterrent to self-enhancement: The search for mechanisms. *Journal of Personality and Social Psychology, 83*, 592-605.

Sedikides, C., & Luke, M. (in press). On when self-enhancement and self-criticism function adaptively and maladaptively. In E. C. Chang (Ed.), *Self-criticism and self-enhancement: Theory, research, and clinical implications*. Washington, DC: APA Books.

Sedikides, C., Rudich, E. A., Gregg, A. P., Kumashiro, M., & Rusbult, C. (2004). Are normal narcissists psychologically healthy?: Self-esteem matters. *Journal of Personality and Social Psychology, 87*, 400-416.

Sedikides, C., & Skowronski, J. A. (1997). The symbolic self in evolutionary context. *Personality and Social Psychology Review, 1*, 80-102.

Sedikides, C., & Skowronski, J. J. (2000). On the evolutionary functions of the symbolic self: The emergence of self-evaluation motives. In A. Tesser, R. Felson, & J. Suls (Eds.), *Psychological perspectives on self and identity* (pp. 91-117). Washington, DC: APA Books.

Sedikides, C., Skowronski, J. J., & Dunbar, R. I. M. (2006). When and why did the self evolve? In M. Schaller, J. A. Simpson, & D. T. Kenrick (Eds.), *Evolution and social psychology: Frontiers in social psychology* (pp. 55-80). New York, NY: Psychology Press

Sedikides, C., & Strube, M. J. (1997). Self-evaluation: To thine own self be good, to thine own self be sure, to thine own self be true, and to thine own self be better. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 29, 209-269). New York, NY: Academic Press.

Silvia, P. J., & Gendolla, G. H. E. (2001). On introspection and self-perception: Does self-focused attention enable accurate self-knowledge? *Review of General Psychology, 5*, 241-269.

Sorrentino, R. M., & Hewitt, E. C. (1984). The uncertainty-reducing properties of achievement tasks revisited. *Journal of Personality and Social Psychology, 47*, 884-889.

Stapel, D. A., & Schwinghammer, S. A. (2004). Defensive social comparison and the constraints of reality. *Social Cognition, 22*, 147-167.

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 21, pp. 261-302). Hillsdale, NJ: Erlbaum.

Swann, W. B., Jr., Rentfrow, P. J., & Guinn, J. (2003). Self-verification: The search for coherence. In M. Leary & J. Tangney (Eds.), *Handbook of self and identity* (pp. 367-383): Guilford, New York.

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

Taylor, S. E., Lerner, J. S., Sherman, D. K., Sage, R. M., & McDowell, N. K. (2003). Portrait of the self-enhancer: Well-adjusted and well-liked or maladjusted and friendless? *Journal of Personality and Social Psychology, 84*, 165-176.

Tesser, A. (1978). Self-generated attitude change. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11, pp. 289-338). New York: Academic Press.

Tesser, A. (2001). On the plasticity of self-defense. *Current Directions in Psychological Science, 10*, 66-69.

Tetlock, P. E., Skitka, L., & Boettger, R. (1989). Social and cognitive strategies for coping with accountability: Conformity, complexity, and bolstering. *Journal of Personality and Social Psychology, 57*, 632-640.

Tice, D. M., Butler, J. L., Muraven, M. B., & Stillwell, A. M. (1995). When modesty prevails: Differential favorability of self-presentation to friends and strangers. *Journal of Personality and Social Psychology, 69*, 1120-1138.

Titchener, E. B. (1912). The schema of introspection. *American Journal of Psychology, 23*, 485-508.

Trope, Y. (1979). Uncertainty-reducing properties of achievement tasks*.* *Journal of Personality and Social Psychology, 37*, 1505-1518.

Trope, Y. (1980). Self-assessment, self-enhancement, and task preference. *Journal of Experimental Social Psychology, 16*, 116-129.

Trope, Y. (1983). Self-assessment in achievement behavior. In J. M. Suls & A. G. Greenwald (Eds.), *Psychological perspectives on the self* (Vol. 2, pp. 93-121). Hillsdale, NJ: Erlbaum.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1130.

Weiner, B. (1985). "Spontaneous" causal thinking. *Psychological Bulletin*, *97*, 74-84.

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

Wilson, T. D., & Dunn, E. W. (2004). Self-knowledge: Its limits, value, and potential for improvement. *Annual Review of Psychology, 55*, 493-518.

Wilson, T. D., Dunn, D. S., Kraft, D., & Lisle, D. J. (1989). Introspection, attitude change, and attitude behavior consistency: The disruptive effects of explaining why we feel the way we do. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 22, pp. 287-343). San Diego, CA: Academic Press.

Wilson, T. D., Kraft, D., & Dunn, D. S. (1989). The disruptive effects of explaining attitudes: The moderating effect of knowledge about the attitude object. *Journal of Experimental Social Psychology, 25*, 379-400.

Wilson, T. D., Lisle, D. J., Schooler, J. W., Hodges, S. D., Klaaren, K. J., LaFleur, S. J. (1993). Introspecting about reasons can reduce post-choice satisfaction. *Personality and Social Psychology Bulletin, 19*, 331-339.

Wosinska, W., Dabul, A. J., Whetstone-Dion, R., & Cialdini, R. B. (1996). Self-presentational responses to success in the organization: The costs and benefits of modesty. *Basic and Applied Social Psychology, 18*, 229-242.

Wundt, W. (1894). *Lectures on human and animal psychology*. (Trans. By J. E. Creighton & E. B. Tichener). New York: Macmillan.

Author Note

Constantine Sedikides, University of Southampton, England, UK; Robert S. Horton, Wabash College; Aiden P. Gregg, University of Southampton, England, UK.

We thank Brandy Carr, Amy Claasen, and Torrey Rieser for their assistance with data collection and coding. Correspondence concerning the manuscript should be addressed to Constantine Sedikides, Center for Research on Self and Identity, School of Psychology, University of Southampton, Highfield Campus, Southampton SO17 1BJ, England, UK. Email: cs2@soton.ac.uk.

Footnotes

[1] An obscure impulse towards pedantry obliges us to specify that “average” here denotes either the mean of a symmetrical distribution or the median of a nonsymmetrical one.

[2] Our research was an expedition into new empirical territory. We were consequently keen to maximize the strength of our key manipulation, and so fashioned it from a mix of reasons-analysis and debiasing elements, each of which was capable of effecting psychological change in its own right. Our chief concern, in the first instance, was to establish that self-enhancement *could* be curtailed in view of its potency and preeminence; hence, developing for an initial “sledgehammer” struck as the most prudent course of action, as well as that most likely to generate a egotism-reducing technique of any practical utility (see General Discussion).

[3] The degrees of freedom in our reasons analyses differ from those reported previously. We were unable to include in these analyses participants (*N* = 11) who chose to take with them their explanatory reasons pages. It is important to note, however, that these 11 participants were distributed across all four conditions of our Target Type X Trait Valence design, with *N*s ranging from 1-4.

[4] Like most personality and social psychologists, we construe narcissism as a normally distributed individual difference, operationalized in terms of relatively high scores on the Narcissistic Personality Inventory (NPI; Raskin & Hall, 1979).

[5] Whereas our experimental manipulation of explanatory introspection instructed participants to consider reasons why they might or might not possess positive or negative traits, explanatory introspection in everyday life, especially when self-esteem is low, may primarily involve people considering reasons why they do have negative traits and why they do not have positive ones. Thus, although the introspection engaged in would still be explanatory (as opposed to, say, descriptive) some of its parameters would vary. We leave it to future research to tease out the differential effects of the various possible forms of explanatory introspection.

Table 1

Trait Self-Descriptiveness, Valence, and Importance Ratings in the Pretest

I. Positive Traits

Trait	*Importance*	*Valence*	*Self-Descriptiveness*
Friendly	9.88	9.33	9.10
Honest*	10.29	10.02	9.21
Independent	9.05	9.13	8.38
Interesting	9.66	9.79	8.97
Kind*	9.80	9.64	9.08
Modest	7.84	7.38	7.13
Non-conformist	6.97	6.77	6.12
Non-judgmental	8.93	8.56	6.52
Organized	8.51	8.93	7.67
Patient	8.44	7.77	6.30
Secure	9.39	9.36	7.13
Trustworthy*	10.43	10.10	9.49

II. Negative Traits

Trait	*Importance*	*Valence*	*Self-Descriptiveness*
Conformist	6.58	6.90	4.45
Dependent	6.97	8.28	5.30
Dishonest*	9.66	9.93	2.20
Disorganized	7.36	8.37	3.72
Immodest	7.41	7.11	4.41
Impatient	7.15	7.49	5.29
Insecure	7.26	8.14	4.75
Judgmental	7.77	8.44	4.66
Unfriendly	9.14	9.03	2.93
Uninteresting	8.11	9.90	2.18
Unkind*	9.11	9.36	2.43
Untrustworthy*	9.98	9.85	1.97

Note 1: Asterisks indicate traits selected for use in the experiments.
Note 2: For positive traits, higher numbers indicate more trait self-descriptiveness, more importance to have the trait, and more trait positivity. For negative traits, higher numbers indicate more trait self-descriptiveness, more importance not to have the trait, and more trait negativity.

Table 2

Self-Descriptiveness Means (and SDs) as a Function of Introspection Target Type, Activity Type, Cognitive Activity, and Trait Valence in Experiment 2

I. SELF AS INTROSPECTION TARGET

A. *Written*

	Explanatory Introspection	Descriptive Introspection
Positive	10.80 (2.22)	12.90 (1.20)
Negative	5.70 (2.03)	4.13 (2.46)

B. *Mental*

	Explanatory Introspection	Descriptive Introspection
Positive	12.57 (1.26)	12.37 (1.59)
Negative	3.53 (1.87)	3.70 (1.53)

II. OTHER AS INTROSPECTION TARGET

A. *Written*

	Explanatory Introspection	Descriptive Introspection
Positive	11.77 (1.10)	12.20 (1.42)
Negative	2.73 (.81)	3.53 (1.47)

B. *Mental*

	Explanatory Introspection	Descriptive Introspection
Positive	12.27 (1.11)	12.83 (.81)
Negative	3.63 (.85)	4.40 (1.62)

Table 3

Self-Descriptiveness Means (and SDs) for Orthogonal Contrasts in Experiment 2

	Introspection Type		
Trait Valence	Explanatory	Descriptive	Control
Positive	10.80 (2.22)	12.90 (1.20)	12.63 (1.27)
Negative	5.70 (2.03)	4.13 (2.46)	3.73 (1.60)

Table 4

Time 1 Certainty, Time 2 Self-Descriptiveness, and Time 2 Certainty Means (and SDs) as a Function of Introspection Type in Experiment 3

		Introspection Type	
	Explanatory	Descriptive	Control
Time 1 Certainty	12.03 (2.09)	13.03 (1.59)	13.49 (1.49)
Time 2 Self-Descriptiveness	12.31 (1.84)	13.14 (.99)	13.62 (.77)
Time 2 Certainty	12.91 (1.83)	13.37 (1.43)	13.70 (1.51)