Peer-Reviewed Letter

IS IDENTITY *PER SE* IRRELEVANT? A CONTRARIAN VIEW OF SELF-VERIFICATION EFFECTS

Aiden P. Gregg*

Self-verification theory (SVT) posits that people who hold negative self-views, such as depressive patients, ironically strive to verify that these self-views are correct, by actively seeking out critical feedback or interaction partners who evaluate them unfavorably. Such verification strivings are allegedly directed towards maximizing subjective perceptions of prediction and control. Nonetheless, verification strivings are also alleged to stabilize maladaptive self-perceptions, and thereby hindering therapeutic recovery. Despite the widespread acceptance of SVT, I contend that the evidence for it is weak and circumstantial. In particular, I contend that that most or all major findings cited in support of SVT can be more economically explained in terms of raison oblige theory (ROT). ROT posits that people with negative self-views solicit critical feedback, not because they want it, but because they their self-view inclines them regard it as probative, a necessary condition for considering it worth obtaining. Relevant findings are reviewed and reinterpreted with an emphasis on depression, and some new empirical data reported. Depression and Anxiety 26:E49–E59, 2009.

Key words: *self-verification; identity; depression; feedback-seeking; rationality*

"I dislike arguments of any kind. They are always vulgar, and often convincing."

— Lady Brackwell, in Oscar Wilde's *The Importance of Being Earnest*

I would like to thank Professor Uhde for inviting me to comment upon the recently published article by[1] entitled *Self-verification and depression in abused women*. Before the article's publication, I had the privilege of being asked to review it. In my review, I observed that the research reported in the article, whatever its other merits or demerits, relied upon a set of fundamental assumptions, namely, those underlying *self-verification theory* (SVT).[2–4] In this article, I wish to challenge those assumptions. This is an ambitious undertaking, given that SVT currently enjoys the status of conventional wisdom within the field social and personality psychology. Moreover, most of what I say will amount to a conceptual critique of existing research, although I shall present some preliminary findings to bolster my case. In addition, although I will review a range of findings, I shall focus on depression, and will conclude by specifically addressing the research reported by.[1]

THE ESSENCE OF SVT

SVT proposes that people habitually seek to *confirm their existing self-views*. The alleged reason? A *coherent* sense of oneself is vital for effective psychological and interpersonal functioning.[2–5] In particular, without a stable self-concept, people's perceptions of their capacity to *predict* and *control* their own lives, and their interactions with others, would be fatally undermined.[6,7] As a result, people, needing to believe that they are who they think they are, *strive to verify* that

School of Psychology, University of Southampton, Southampton, United Kingdom

*Correspondence to: Aiden P. Gregg, Shackleton Building, Southampton, Hampshire SO17 1BJ, UK.
E-mail: aiden@soton.ac.uk

Received for publication 13 September 2007; Accepted 28 September 2007

DOI 10.1002/da.20428
Published online 9 December 2008 in Wiley InterScience (www.interscience.wiley.com).

they are who they think they are.[1] They do this principally by embracing information consistent with their existing self-views, and disdaining information inconsistent with those self-views. Importantly, it should not matter, for the purposes of safeguarding psychic coherence, whether one's self-views are positive or negative, true or false: what should matter, rather, is that they are *consistently held*. In other words, SVT makes the bold claim that *identity matters per se*, above and beyond the valence or accuracy of its content.

THE IMPLICATIONS OF SVT

Now, if SVT is *true*, then it has an unsettling implication: people who hold negative self-views will at some level *want* to believe that those negative self-views are true. And, of course, wanting to believe that they are true, they will be averse to disconfirming them. Such aversion will then impede the success of any therapeutic intervention aimed at banishing those negative self-views. For example, chronic depression would be hard to alleviate, not *only* because the condition itself is functionally or biochemically stubborn, but *also* because those who suffer from it, having regarded themselves as depressed for so long, would irrationally *seek to keep seeing themselves as depressed*, for fear their long-standing identity might be undermined. Accordingly, people with chronic depression would solicit feedback that verifies rather than refutes their negative self-view, thereby hindering their recovery.

Now, although this implication of SVT is unsettling, it at least promises to shed light on the seemingly irrational persistence of many negative self-views. For example, people suffering from low self-esteem, whether as a standalone condition or as a symptom of depression, tend to regard themselves as worthless and undeserving, even though their objective virtues and achievements may strongly indicate otherwise.[8] Such unwarranted self-negativity is prima facie puzzling. However, SVT contends that people cling to their unflattering images of self because those are the only self-images they have, and without them they would be lost. Simply put, people endure the frying pan of self-reproach to escape the flames of self-uncertainty.

However, if SVT is *false*, then it would be positing an illusory impediment to recovery from clinical conditions that, like depression, entail a negative self-view. Therapeutic applications based on SVT, or those taking account of it, would therefore be misdirected, conceivably to the detriment of their efficacy. Moreover, the message of SVT would be unduly pessimistic—a hindrance to people whose capacity to hope is already impaired, and whose positive expectations need to be mobilized. Clearly, then, it matters a great deal whether SVT is true or false. Consequently, its theoretical and empirical bases merit scrutiny.

THE EVIDENCE FOR SVT

On the face of it, the evidence cited in favor SVT looks robust, for four reasons. First, the evidence is predominantly *behavioral* in nature. Rather than merely consisting, say, of subjective reports about which type of self-related feedback or interaction partners people prefer, it also consists of their objective choices for one type of self-related feedback or interaction partner over another.[9–11] Second, the evidence is *convergent*. Findings from a variety of domains—from patterns of attention([12]: Study 1) to martial intimacy[13]—seemingly support SVT. Third, the evidence consists, not only of predicted main effects, but also of predicted *moderator effects*.[14] This means that any rival theory has more explanatory work to do. Finally, the evidence consists of several findings that reliably replicate[4,15]). Hence, SVT does not merely capitalize upon statistical flukes.

Space constraints preclude an exhaustive summary of the evidence for SVT. Fortunately, some of the best evidence derives from studies purporting to show that depressed people strive to verify their self-views—the claim of potentially greatest interest to readers of this journal. Hence, these studies will be the focus of my initial exposition and subsequent critique. The key articles are.[16–18]

EMPIRICAL STUDIES OF SELF-VERIFICATION EFFECTS IN DEPRESSED PEOPLE

Suppose people with a positive self-view (e.g., joyful, optimistic, and self-accepting) were to characteristically solicit information consistent with their self-views (i.e., flattering information). Would that suggest the operation of a motive to verify their existing identity? No. This reason is that such behavior would equally suggest the operation of another well-known and potent motive: to self-enhance.[19,20] However, if people with a negative self-view (e.g., gloomy, pessimistic, and self-critical) solicited information consistent with their self-views, this would not be the case. Rather, such a pattern of solicitation would imply that the motive to self-enhance could even be *subverted* by a more powerful motive.

So what does the relevant research show? Indeed, unlike cheerful people, depressed people are

[1]SVT draws a distinction between *epistemic* self-verification (ESV), which aims at preserving intrapsychic coherence, and *pragmatic* self-verification (PSV), which aims at keeping social interactions running smoothly. In this letter, I concern myself primarily with ESV. The reason is that ESV, unlike PSV, is hypothesized to have the reinforcement of identity as a *direct* goal. That is, although PSV may, through the feedback it elicits, *result* in identity being reinforced, its hypothesized direct goal is the fostering of interpersonal harmony. In this regard, PSV seems like a subspecies of the evolutionarily adaptive *need to belong* (Baumeister & Leary, [5]), and its connection to identity may be accidental.

more likely to opt to meet evaluators who rate them *un*favorably as opposed to favorably([16]: Experiment 1). Furthermore, depressed people, when offered the choice between reading a favorable or unfavorable assessment of their personality, mainly opt for the latter, whereas cheerful people in contrast, mainly opt for the former.[18] Finally, depressed people are prone to ask questions that typically yield more critical and less flattering feedback, both from alleged observers actual roommates and alleged observers[16]: Experiments 3 and 4). In sum, depressed people reliably act in ways that lead to them receiving information in keeping with their negative self-views.

There are also indications that depressed people *actively* solicit negative information. For example, depressed people not only opt to interact with evaluators who rate them unfavorably as opposed to favorably, they also do so as an alternative to participating in an unrelated experiment[17]: Experiment 1). In addition, depressed people are more apt to solicit information about specific weaknesses relative to specific strengths after receiving generally flattering feedback; in contrast, cheerful people are more apt to solicit information about specific strengths relative to specific weakness after receiving generally critical feedback[17]: Experiment 2). In sum, depressed people flee flattery and court criticism, even when they need not do so.

Further grounds for postulating a motive to self-verify are provided by self-reported ratings. For example, depressed people, relative to cheerful people, ideally like to viewed less positively (although never actually negatively) by good friends and dating partners[16]: Experiment 2). In addition, whereas cheerful people report a relatively greater desire to interact with favorable evaluators, depressed people report a relatively greater desire to interact with unfavorable evaluators (although the desire to interact rarely dips below moderate;[17]: Experiment 1]. Complementing these results, depressed people report desiring to read an unfavorable personality assessment more than a favorable one, whereas cheerful people report desiring to read an favorable personality assessment more than an unfavorable one (although desire to read both remains generally high).[18] In sum, the pattern of depressed people's self-reported desire for feedback matches the pattern of their actual feedback-seeking.

EMPIRICAL STUDIES OF MODERATORS OF SELF-VERIFICATION EFFECTS

Key moderators of self-verification effects have only been explored outside the domain of depression. Nonetheless, they deserve mention, because they illustrate the scope of SVT, and indicate what additional findings require alternative explanation.

First, self-verification effects are *time-dependent*. In particular, when people with negative self-views make a snap judgment as to whether to interact with an evaluator who views them either favorably or unfavorably, they more often opt to interact with the *former*. Only when they have ample time for reflection do they more often opt to interact with the latter ([14]: Experiment 3). In addition, when under time pressure, people with negative self-views report being keener to interact with an evaluator who views them *favorably* than one who views them unfavorably; this effect is attenuated, as usual, only in the absence of time pressure[14]: Experiment 2; see also[22]: Experiment 4). The standard interpretation of these findings is that self-verification strivings, unlike self-enhancement strivings, are mediated by relatively sophisticated cognitive operations (i.e., involving the comparison of feedback content with self-representations). When time is short, these cognitive operations are undermined, and self-verification effects diminish.

Second, self-verification effects depend on the *importance* and *certainty* of self-views. For example, students with negative self-views, just like those with positive self-views, are more committed to living with their roommates to the extent that those roommates *share* their (negative or positive) self-views, but only when they hold those self-views with certainty and regard them as important.[23] This finding recalls another, in which people, even when their self-view is negative, report greater intimacy with partners who see them as they see themselves, but only when their partners are spouses, not dates[24]: see also[11,21]). The standard interpretation of these findings is that, when identity is more strongly linked to self-views and relationships, it matters more. Hence, verification strivings are potentiated.

A CRITIQUE OF SVT

I will now argue that, despite all the above findings, there are insufficient grounds for concluding that, in order to safeguard psychic coherence, people *want* to self-verify—that is, desire to confirm that they are the type of person they already think they are. Note that I will *not* be contesting the claim that people *act* in ways that result in their identities being reinforced; the empirical evidence, reviewed above, clearly indicates that they do. Most notably, that evidence indicates that even people with highly negative self-views act in such ways. Indeed, SVT research is to be commended for highlighting potentially maladaptive forms of self-confirmation.[25] It is also to be commended for empirically exploring a bold and counterintuitive account of human motivation. Nonetheless, SVT remains, like all scientific theories that make substantial claims, open to perpetual disconfirmation. The enduring validity of SVT is established, not merely by accumulating inductive evidence consistent with its predictions, but by repeatedly testing whether it

accounts for empirical findings better than plausible rival theories.[26] In this regard, I shall shortly outline an alternative to SVT, called *raison oblige theory* (ROT) for reasons that will become apparent. In brief, ROT proposes that most or all of the findings cited in support of SVT are a product, not of any verification strivings, but of the obligations that everyday rationality imposes on people.

BEHAVIOR AND MOTIVE: DISTINCT AND NOT ISOMORPHIC

Before outlining ROT, two points need to be emphasized. First, behavior and motive are different things. Second, behavior need not imply a corresponding motive.

It might seem obvious that behavior and motive are distinct entities; indeed, it *is* obvious. To spell it out, behavior is what someone physically does, motive is what mentally makes them do it. However, suppose it is claimed that a person *self-verified*. Does this *merely* mean that the person acted so as to solicit information consistent with their self-view, or does it *additionally* mean that they acted in this way under the influence of the motive to self-verify? The ambiguity of the verb (and related terms) leaves room for equivocation. It also creates the possibility that self-verification, as a behavioral effect that might be variously explained, could be inadvertently construed as a behavioral effect with only one explanation, namely a motive to stabilize identity. Accordingly, the distinction between behavior and motive must be kept clear.

In general, it is perilous to postulate motives that map on perfectly to particular classes of behavior. For example, people occasionally do things that are self-defeating (e.g., abusing drugs) or self-destructive (e.g., committing suicide). Observing such oddities, psychoanalytic thinkers have been prone to infer that isomorphic impulses must lie behind such acts—the "death instinct" being the most well-known example ([27]; see also[28]). However, empirical investigations suggest that, when people act against their own best interests, it is not because they *want* to do so, but rather because they wish to pursue *other* incompatible interests, typically directed at the immediate alleviation of negative affect.[29] Hence, care must be taken not to rashly conclude that *one* motive, specific to a class of behaviors, is *the* explanation for those behaviors, when a host of rival motives remain viable. To their credit, Swann and colleagues—perhaps prompted by earlier criticisms[30]—have attempted to rule out at least some explanations for why people with negative self-views solicit critical (or avoid flattering) information.[11,21,24] It now seems unlikely that the effect can be put down, in any comprehensive way, to people seeking to modify their partner's perception of them, searching for information to help them improve themselves, or looking for like-minded others to validate their attitudes. However, the perennial difficulty is that there could always be some *additional* explanation for the effect that has not yet been tested (compare research[27] seeking to rule out egoistic motives for alleged altruistic helping). Indeed, this is precisely what I contend.

There is a further reason why inferring isomorphic motives from behavior is a tricky business: people sometimes voluntarily do things that they do *not* want to do, and sometimes voluntarily *refrain* from doing things that they do want to do. An example of the former would be reluctantly giving up leisure time to perform administrative duties; an example of the latter would be resisting the urge to invent qualifications in order to secure a coveted job. If one accepts that people sometimes disregard or override their desires, then it immediately follows that one cannot automatically infer, simply because someone has done something, that they wanted to do it. Rather one would have to concede that behavior, as an index of wanting, is intrinsically fallible. Of course, it might still be the case that, most of the time, people indeed do what they want, so that, generally speaking, behavior *is* a good baseline indicator of wanting. Nonetheless, there might also be the case that, under particular circumstances, behavior is particularly *undiagnostic* of wanting. Indeed, I contend below that most of the behavioral evidence cited in support of SVT has been gathered under precisely such circumstances.

RAISON OBLIGE THEORY

What does all the above have to do with opting for one type of feedback or interaction partner over another? Consider again the two instances given above of voluntary yet uncongenial acts. The reason that someone would give up leisure time to perform administrative duties is that they are *obliged* to do so; and the reason that they would resist the urge to invent qualifications is that they are *not entitled* to do so. Such obligations and absences of entitlement, familiar to everyone, are *normative* phenomena. Here, the norms in question, which guide behavior, are moral in nature. However, norms need not only be moral: they can also be *rational*. Just as people are not at liberty to act however they want, so people are not at liberty to believe whatever they want. In particular, they are obliged to believe some propositions they would rather *not* believe, and not entitled to believe others propositions that they *would* rather believe. I will first illustrate this abstract point with a simple if fanciful example, and then extend the reasoning to the more sober subject of depression.

THE UGLY DUCKLING'S DILEMMA

Suppose I am a duckling with a negative self-view. I earnestly believe, rightly or wrongly, that I am ugly.

Further suppose that, thanks to my enhancement strivings, I nonetheless yearn to be a beautiful swan. Alas, despite the strength of my strivings, I feel certain that I can never *be* a beautiful swan.[2]

Now imagine that two other ducklings, A and B, share my pond. According to reliable sources, Duckling A has a favorable view of my appearance, regarding me as beautiful, whereas Duckling B has an unfavorable view of my appearance, regarding me as ugly. The opportunity now arises for me to interact with either Duckling A or Duckling B. Which one should I choose to interact with and why?

Suppose I opt to interact with Duckling B. (Moreover, to do justice to the empirical literature, also suppose that I (a) claim that I want to do so, (b) would do so even when the opportunity arose to do something else, and (c) am not doing so either to improve my appearance, change anyone's opinion of me, or validate my attitudes.) According to SVT, my choice of interaction partner would constitute evidence that I *wanted* to believe that I was ugly. However, ROT suggests an alternative interpretation. Given that I earnestly believe that I am ugly, I am (a) obliged to believe that Duckling B's critical view of my appearance is true, and (b) not entitled to believe the Duckling's A flattering view of my appearance is true. In other words, my self-view entails that I believe Duckling B is right about me and Duckling A is wrong about me. But if I truly believe this, then it follows that what Duckling A believes about me, despite being congenial, *cannot* strike me as probative, because *any* proposition I regard as false cannot strike me as probative. On the other hand, what Duckling B believes about me, despite being uncongenial, *may* strike me as probative, because any proposition I regard as true at least satisfies a *necessary condition* for striking me as probative. Consequently, if information about myself is what I seek, then I have some grounds for interacting with Duckling B but no *grounds* for interacting with Duckling A. True, I do have a *motive* to interact with Duckling A over Duckling B: the former praises me whereas the latter deprecates me. However, because I am a rational as well as a willful being, the choices I make are based, not only on what I ideally *want* to believe, but also on what I legitimately *can* believe—indeed, *largely* on the latter. Hence, I choose to interact with Duckling B, not Duckling A. The line of reasoning lies at the heart of ROT.

Given the thrust of much recent research on the self,[31] one might for forgiven for concluding that rationality is dwarfed by the motive to self-enhance. However, this would be a gross error. In fact, rationality is pervasive and motives merely qualify it. Were it not so, grandiose delusions would be common, and realistic self-assessments rare. Yet the pervasive impact of rationality is easily overlooked precisely because it is part of the taken-for-granted architecture of everyday cognition.[32] Consider one well-replicated effect attributed to self-enhancement: the *above-average effect*.[33] In general, people evaluate themselves as superior than their peers on common personality traits.[34] However, the effect only emerges if the traits in question are sufficiently ambiguous to permit a generous latitude of interpretation.[35] Reality constrains the motive to self-enhance: respondents may *want* to believe they compare favourably to peers on personality traits, but they *cannot* when those traits are well-defined. In other words, *raison* (like *noblesse*) *oblige*: to be reasonable is to accept that one's beliefs *must* be based on grounds, not motives. Credibility generally trumps desirability.

WHY ROT EXPLAINS SELF-VERIFICATION EFFECTS

Depressed people find themselves in the same type of dilemma as the hypothetical duckling above. They may *wish* to see themselves as joyful, optimistic, and self-accepting, but they are only *able* to see themselves as gloomy, pessimistic, and self-critical. Hence, they are obliged to regard feedback about themselves as probative only if it accords with their negative self-view. Flattering feedback about themselves strikes them as eminently desirable but unfortunately incredible, whereas critical feedback strikes them as unfortunately credible but eminently undesirable. As a result, there is *no need* to posit a self-verification motive to explain why depressed people would choose the latter over the former (or interaction partners who might provide one over the other). Choosing flattering feedback would seem pointless to them because the information offered would seem wrong; yet choosing the former *might* seem worthwhile to them because the information offered would seem right.[3] The basic point is that depressed people, being constrained by everyday reason, tend to opt for feedback they think they *merit* rather than for the feedback they most want to be true. And although depressed people may hold irrationally negative self-views, once they assume those negative self-views are true, their appraisals of feedback duly come to accord with their assumptions. So it is not, as SVT maintains, that depressed self-views solicit critical feedback out of a *psychological desire* to confirm existing self-views. Rather, they solicit critical feedback out of

[2]Don't worry: the story has a happy ending. Moreover, when the ugly duckling becomes a beautiful swan, no identity crisis ensues, unlike SVT would predict.

[3]If only feedback featuring information consistent with one's self-view can be regarded as probative and justified, then effects on attention and memory attributed to verification strivings would also admit of alternative explanation (Swann & Read, [43]: Experiment 1 and 3). In particular, people with negative self-views would both attend to and recall critical feedback better (just like people with positive self-views would tend to recall flattering feedback better) because they would perceive that feedback to be relatively probative and justified.

rational obligation to honor feedback in keeping with their existing self-views.

SOME QUESTIONABLE EVIDENCE FOR SVT

Does ROT fit the empirical facts? I argue that it does. Moreover, I argue that some findings cited in support of SVT either support ROT better or even call SVT into question.

First, self-verification effects are consistently reported as being driven by cognitive rather than motivational factors.[36–38] For example[16] found that, in a sample of cheerful and depressed people, ratings of feedback credibility and self-descriptiveness intercorrelated highly, and when amalgamated into a single index, predicted interest in interacting with evaluators (Experiment 1). In addition, both[17] and [18] found that, whereas cheerful people regarded flattering feedback as more accurate than critical feedback, depressed people regarded critical feedback as more accurate than positive feedback. They also found that, whereas feedback desirability did not predict feedback or interaction partner choice, perceived feedback accuracy did (see also[37]). Finally, in normal samples,[39] found that perceived feedback accuracy fully mediated feedback choice, while[[9]: Experiment 3] found that feedback consistent with self-views was regarded as especially informative.

The standard SVT interpretation of these findings runs as follows. In their quest to self-verify, people—even depressed people with negative self-views—are motivationally biased towards seeing self-descriptive feedback as accurate, credible, and informative. Indeed, so strong is this bias, that it entirely outweighs the bias towards seeing flattering feedback as accurate, credible, and informative. The alternative ROT interpretation runs as follows. People have no motivational bias toward seeing self-descriptive feedback as accurate, credible, and informative: they simply *do*, in virtue of rationally extrapolating, on the basis of the self-view that they earnestly hold, that feedback consistent with their self-view is liable to be true and feedback inconsistent with it likely to be false.

Thus, both SVT and ROT are consistent with the relevant data. However, whereas SVT neglects to take into account everyday rationality and postulates an extra motive to account for the obtained effects, ROT endeavors to efficiently explain the obtained effects solely on the basis of everyday rationality. Thus, if Occam's razor were to be invoked, ROT would be preferred.

Second, SVT predicts that verification strivings rely on particular cognitive operations (i.e., comparing feedback content with self-representations). However, it should *also* predict that verification strivings influence affective state: strivings, after all, can be agreeably satisfied or disagreeably frustrated. In particular, a person with a negative self-view should be *less* emotionally disturbed by critical feedback or by interaction partners who view them unfavourably. This is because whereas critical feedback would challenge *both* the enhancement and verification strivings of a person with a positive self-view, it would *only* challenge the enhancement strivings of a person with a negative self-view while *also* supporting their verification strivings. Thus, whereas people with a positive self-view would have both their ego *and* their identity threatened—a potentially dismaying experience on all counts—people with a negative self-view would *only* have their ego threatened while also having their identity reinforced—but a potentially dismaying experience on one count, but a reassuring one on another. A converse argument could also be put forward to predict that people with negative self-views should be less emotionally buoyed by flattering feedback.

Yet what evidence exists suggests that people with positive and negative self-views do not differ in how disturbed they are by criticism or in how buoyed they are by flattery.[40] In particular,[37] found that people with positive and negative self-views generally did not differ in their emotional reactions to flattering and critical feedback: both ended up more moody, depressed, anxious, and hostile after receiving the latter than after receiving the former. In addition, both liked the source of flattering feedback but disliked the source of critical feedback (although people with negative self-view did report being significantly more *attracted* the source of negative feedback). Is it not odd that having one's identity reinforced or undermined does *not* moderate how a person feels? Given these null findings, can identity per se really be the critical factor in maintaining clinical conditions such as depression and anxiety? One unpublished study[41] did find that at least when flattering feedback came from a credible source, people with negative self-views were more anxious upon receiving it. However, it is moot point whether that anxiety was occasioned by frustrated verification strivings. Might it not have derived simply from the understandably distressing realization that they were not entitled to believe the flattering feedback?

WHY ROT EXPLAINS MODERATORS OF SELF-VERIFICATION EFFECTS

ROT also accounts nicely for why people with negative self-views choose to interact with evaluators who view them unfavorably when they can make a unhurried choice, but evaluators who view them favorably when they must make an immediate choice.[14] Quite simply, time pressure imposes a cognitive load that disables the explicit cognitive processes required to differentiate truth from falsity, but it leaves intact the implicit cognitive processes required to differentiate positivity from negativity, (see [42] for a general cognitive model). Hence, people with negative self-views spontaneously seek the

"nice" evaluator but spurn the "nasty" one: any gloomy reflections about their rational entitlements and obligations are nipped in the cognitive bud. The same mechanism is likely to at least partly account for some other conceptually similar research findings.[43,44]

ROT also accounts nicely for why self-verification effects occur only when self-views are held with certainty and seen as important.[23] The certainty and importance of a self-view reflect the earnestness with which it is believed. The more earnestly a self-view is believed, the more information compatible with it will be deemed true and worth obtaining, and the more information incompatible with it will be deemed false and worth ignoring. Hence, the more certain and important a negative self-view, the more one should opt for critical over flattering feedback, and the less certain and important a negative self-view, the more one should opt for flattering over critical feedback. This would explain why depressed people, whose negative self-views are more ingrained, show stronger self-verification effects than people with merely low self-esteem[16]: Experiments 1 and 2;[18], as well as why self-verification effects are stronger for well-elaborated and schematic traits.[45,46]

TWO LOOSE ENDS: PARTNER PERCEPTIVENESS AND SPONTANEOUS VERBALIZATIONS

My critique so far shows that ROT can plausibly and efficiently account for most lines of evidence cited in favor of SVT. However, it would be remiss not to cover two further lines of evidence cited.

First, take the following pair of findings: (a) spouses are more committed to and intimate with each another to the extent that their partners see them as they see themselves[24] and (b) spouses seek to disabuse their partners of views of themselves that they do not share.[13] Neither finding actually offers strong evidence for verification strivings. For example, it follows almost by definition that I will rate a partner who sees me as I see myself as more intimate with me. Moreover, if a partner who sees me more negatively also treats me worse—a plausible correlation—I might, through cognitive dissonance, escalate my commitment to the relationship.[47] In addition, if a partner fails to see me as I see myself, then it will subjectively seem to me that my partner is mistaken, and thereby become my perceived responsibility to correct ll, it's wrong to let others simply labour under a delusion, especially if one happens to be married to them.[4]

However, an ancillary finding in these studies seems to contradict ROT. Ratings of *partners' perceptiveness* were not associated with either marital commitment or intimacy.[24] But doesn't ROT predict that (ultimately long-term) interaction partners will be (consistently) chosen depending on the probative value of feedback they offer, and shouldn't this vary as a function of that partner's rated perceptiveness? Not necessarily. As it turns out, the item assessing partner perceptiveness in these studies pertained to their perceptiveness *generally* rather than within the relationship specifically.[24] Furthermore, it would be as much a problem for SVT as for ROT if partners' rated perceptiveness within the relationship did *not* predict levels of commitment or intimacy. After all, if verification strivings are to be satisfied, someone must believe that someone else sees them as they see themselves, that is, accurately and perceptively.

Second,[11] had participants with positive and negative self-views opt for an interaction partner who evaluated them either favorably or unfavorably, and then verbalize their reasons for having done so. Participants' verbalizations were subsequently coded by four naïve raters into different categories. The final set of categories included ESV and PSV (see Footnote 1). A second group of raters then recoded the verbalizations in terms of these categories. Results suggested that, whereas both enhancement and verification strivings prompted people with positive self-views to seek partners who evaluated them favorably, only verification strivings prompted people with negative self-views to seek partners who evaluated them unfavorably. Given the seemingly bottom-up nature of this investigation, the study would seem to furnish reasonably good evidence for verification strivings.

However, the manner in which categories were derived is unclear from the article. From a footnote[11] p. 394), it looks possible that no more than one rater initially arrived at the epistemic self-verification category independently, and no information was given about how prominently particular categories featured.[5] Furthermore, whether a category generated by the rater actually corresponded to the category specified by SVT was left up to the authors to decide subjectively. Finally, it is not stated what criterion the authors used to rule in or rule out alternative candidate categories.

Admittedly, the pattern of subsequent ratings did accord neatly with what SVT predicts. However, some niggling doubts remain. In particular, one wonders whether, having been provided with the category ESV, raters tended to shoehorn some otherwise difficult to classify verbalizations into it. Here is the description of the category in question: "The evaluator put the speaker at ease by confirming his self-view. Explana-

[4]This might be construed as an instance of pragmatic self-verification. But again, what this has to do with striving to verify one's identity is unclear.

[5]See Gregg et al. [17] for a methodology designed to estimate category importance on the basis of frequency and priority with which coded exemplars are collectively mentioned.

tion: The speaker was reassured that he really knew himself because the evaluator confirmed his self-conception." Two features of this category should be noted: (a) it is comparatively complex and abstruse; and (b) it requires raters to infer difficult to observe intrapsychic processes. Both features might conceivably have reduced the validity of the classification. In addition, consider the three verbalization excerpts the authors chose to illustrate the category,[11] p. 401): (a) "[the unfavorable evaluator] better reflects my own view of myself from experience" (b) "…[the unfavorable evaluator] seems more accurate about what I think about myself...**I'd feel more at ease** with someone who…can actively judge me for what I am" and (c) "That's just a very good way to talk about me. Well, I mean, after examining all of this I think [the unfavorable evaluator] pretty much has me pegged." Note that, whereas only one excerpt refers to putting the speaker at ease, all three arguably refer to the probative value of the assessment—just as ROT would predict. Of course, no firm conclusions can be drawn from such thin second-hand evidence. However, it would seem that a more detailed analysis of verbalizations is required to establish that verification strivings rather than rational obligations explain people's choices of interaction partner. This is especially so given that, as documented above, there is a conspicuous absence of evidence that critical feedback puts people with negative self-views "at ease", or that flattering feedback disturbs them.

TWO NEW EMPIRICAL STUDIES

So far, my contribution has been negative, taking issue with the fundamental assumptions of SVT. Now to make a positive contribution, I report two studies whose findings tend to favor ROT over SVT.[48] As an innovation, these studies inquire, not only into how much people with positive and negative self-views regard flattering and critical feedback as either desirable or credible,[37] but also into how much they *want such feedback to be true*.

In Study 1 ($N = 179$), participants (mostly female students) either filled out a booklet in class, or responded to a computer program by themselves (it did not affect the results). They began by completing a standard measure of global self-esteem.[49] Next, they reported their reactions, on seven-point bipolar scales, to hypothetical feedback directed towards their personalities as a whole. The feedback took the form of two sets of four statements—one set designed to be highly flattering, the other highly critical (e.g., "Generally, I consider you to be a rather fine person" versus "As a person, I don't think much of you").

Three classes of reactions to these statements were then assessed. First, *emotional* reactions, involving anticipated levels of sadness and anger (*I would feel* [upset/hurt/annoyed/offended] *by these statements*). Second, *affective-volitional* reactions, involving anticipated levels of appreciation and desire (*I would* like *what these statements say/I would* want *these statements to be* true / *I would* want to interact *with whoever made these statements*). Finally, *cognitive* reactions, involving the perceived rational plausibility of statements (*I would regard these statements as* accurate / *I would find these statements* difficult to believe), and the anticipated degree to which they would undermined identity (*These statements would disturb my sense of who I am*).

Now, SVT would predict that, relative to people with higher self-esteem, people with lower self-esteem would: (a) anticipate being less emotionally perturbed by critical feedback, and more emotionally perturbed by flattering feedback; (b) anticipate liking critical feedback more, wanting it to be true more, and wanting to interact with those providing it more; (c) anticipate liking flattering feedback less, wanting it to be true less, and wanting to interact with those providing it less; and finally (d) anticipate that critical feedback would disrupt their self-concept less, and that flattering feedback would disrupt it more. In contrast, ROT would predict that the above effects would *not* emerge; it would instead merely predict that, relative to people with higher self-esteem, people with lower self-esteem would simply find critical feedback more plausible, and flattering feedback less.

The results failed to support SVT. For example, anticipated emotional reactions to feedback failed to vary with levels of self-esteem; instead, participants universally anticipated that flattering feedback would arouse pleasant emotions and that critical feedback would arouse unpleasant ones. Moreover, directly contradicting SVT, the *lower* participants' self-esteem, the *less* they anticipated liking critical feedback, and the *less* keen they were for it to be true (rs $\approx .2$). Also contrary to SVT, the lower participants' self-esteem, they more they reported that *any* feedback would disrupt their self-concept (rs $\approx .2$), whether it was critical or flattering, see also [18] p. 364). However, as ROT would predict, critical feedback was regarded as more plausible as self-esteem increased, and flattering feedback as more plausible as self-esteem decreased (r's $\approx .4$).

The upshot of Study 1 is that, even if people with negative self-views want to *receive* critical feedback *more* than people with positive self-views do, the former seem *less keen* than the latter for such critical feedback to be *true* (note: *neither* were very keen). In addition, Study 1 suggests that people with negative self-views, relative to people with positive self-views, find critical feedback more convincing. The conjunction of these two finding suggests that people with negative self-views habitually opt for critical over flattering feedback, not because they are *eager to believe* that their self-view coincides with the feedback chosen, but because they feel *rationally obliged* to concur with its content.

Whereas Study 1 dealt with global self-view, Study 2 dealt with specific self-views. All participants ($N = 141$: high school students and their parents) filled out

individual booklets in several large groups. They began by indicating, on four-point bipolar scales, whether they regarded themselves positively or negatively, in the domain of either attractiveness or intelligence. Next, they read a description of a hypothetical interaction with a person called Chris. As part of a psychology study, this "Chris" had allegedly formed an impression of them, written it down, and departed. Participants were told they would now have to decide, on the basis of what Chris had written, whether or not to interact with Chris again. Chris's impression was randomly set to either confirm or disconfirm their previously expressed self-view. More specifically, on the relevant page of the booklet, participants self-view (positive or negative) was restated, Chris's agreement or disagreement with it was indicated, and Chris's impression of them (flattering or critical) was described. Finally, participants were asked the following three questions: (a) *All other things equal, how much would you want to spend* more *time with Chris*? (b) *How much do you want Chris's opinion of you to be* true*?* (c) *How* accurate *is Chris's opinion of you?*

Only a quarter of participants saw themselves as unattractive, and an even smaller fraction saw themselves as unintelligent. Nonetheless, available results were generally in line with those Study 2, insofar as neither self-view correlated significantly with wanting to interact with Chris or wanting his impression to be accurate. SVT, of course, would have predicted that, compared to participants with more positive self-views, participants with more negative self-views would have shown a greater desire to interact with a "critical" Chris and for his impression of them to be true, as well as a lesser desire to interact with a "flattering" Chris and for his impression of them to be true. However, the *only* correlational finding to emerge was, as in Study 1, that participants with more negative self-views found critical feedback more plausible, and flattering feedback less plausible (r's $\approx .5$). Apart from that, all participants self-enhancingly reported wanting Chris impression to be true when flattering and false when critical. Hence, the findings of Study 1 replicate and generalize, thereby favoring ROT over SVT.

Admittedly, both these studies have limitations. Participants reported how they would respond to feedback as opposed to actually responding to it; and the mediating role for feedback plausibility was inferred from circumstantial evidence as opposed to being directly demonstrated. However, in the defense of the studies, two points might be made. First, the onus is on the skeptic to explain why imaginative introspection should be utterly blind to self-verification strivings that are alleged to cause such powerful behavioral effects (and see[16,39] for methodologically similar investigations). Second, the evidence yielded by these studies is arguably no more circumstantial than that yielded by classic self-verification studies, in which self-verification strivings are neither directly measured or manipulated.

IMPLICATIONS FOR Pineles et al. [1]

It is a prima facie puzzle why abused women, who tend to be depressed,[50] are often so reluctant to leave their abusive male partner. One explanation might be that, having originally freely entered into a relationship, and then invested a lot effort trying to improve it once it turned abusive, they are motivated to rationalize their choices and efforts by concluding that the relationship is worth pursuing, even when it is not.[47,51] However, SVT suggests another explanation: abused women stay in abusive relationships because they want to verify the pre-existing negative self-view that goes along with it. To put it baldly, they would rather get psychologically humiliated and physically hurt than abandon the negative self-views that afford them a sense of predictability and control.

In order to muster evidence for this hypothesis,[1] had groups of abused and nonabused women rate how positively two people would rate specific aspects of their personality: themselves, and a student clinician. Next, both groups of women ranked their relative preference for feedback about each of these aspects, ostensibly to be provided by the clinician. To estimate verification strivings, these rankings were correlated within-subject with the *absolute* differences between the women's self-ratings on each aspect and the ratings the women anticipated the clinician would give them. Greater congruence was deemed evidence of stronger verification strivings, lesser congruence of weaker verification strivings.

As it turned out, the correlation obtained was small ($r = .13$), and did not differ across both groups of women. However, the logic behind the inference that this correlation reflects verification strivings is suspect on multiple grounds.

First, although the use of absolute values ensured that differences in self and anticipated clinician ratings contributed (inversely) to estimates of congruence regardless of the direction of those differences, it could still have been the case that, say, most anticipated clinician ratings exceeded most or all self-ratings, such that women additionally preferring feedback about aspects of personality when the disparities were smaller. If so, then all estimates of congruence would potentially have been contaminated by enhancement strivings (because women would have preferred feedback on those aspects of personality when they did *not* rate themselves worse than the clinician did).

Second, and relatedly, it is not clear what the women's estimates of clinician's ratings were supposed to represent. Were they the ratings that the women would ideally desire, those that they would absolutely dread, or those they would expect a naïve young student clinician to have? In the absence of any definite answer to this question, it is impossible to interpret what the absolute differential between the women's self-rating and the anticipated clinician rating means.

By extension, it is also impossible to interpret what any correlations with this absolute differential mean.

Third—and most relevant to the thrust of this article—the assumption that the congruence between the women's and the anticipated clinician's ratings would solely or even largely a function of verification strivings is untenable. All ratings would have been informed, not only by what the women *wanted* to believe, but also by what they were *obliged* to believe. The authors, I submit, like many other self-verification researches, failed to take adequate account of the degree to which ratings are constrained by everyday rationality.

Pineles et al.[1] also looked at whether abused and nonabused women differed in terms of the correlations between their ranked preferences for feedback and the *signed* differences between their self-ratings on each aspect and the ratings they anticipated the clinician would give them. Here, patterns reminiscent of self-verification effects emerged: whereas nonabused women showed a preference for feedback about aspects of personality where the clinician evaluated them more favorably than they evaluated themselves, abused women did not. Moreover, these effects were statistically mediated by levels of depression. Again, however, these findings cannot be deemed evidence of verification strivings. The possibility that women's ratings could be a function of something other than motive was not even considered, much less eliminated.

CLOSING COMMENTS

In this article, I have attempted to make the case that the existing evidence for self-verification strivings—the desire to confirm that one is who one already believes oneself to be—is much weaker than is commonly asserted. Indeed, the possibility remains open that people *never* strive to bolster their identity per se. Instead, people with negative self-views—including depressives and abused women—may seek out critical feedback and disdain positive feedback, not because they *want* the former to be true and the latter false, but rather because their reason tells that they are *obliged* to believe the former and *not entitled* to believe the latter. The premises from which people with which negative self-views begin, of course, are often tragically unfounded in themselves, but the inferences they deduce from it are not. The real challenge for depression researchers may be, not to curb depressed people's desire to verify that they are gloomy, pessimistic, and self-critical, but instead to find effective ways of convincing them that they can be joyful, optimistic, and self-accepting.

REFERENCES

1. Pineles SL, Mineka S, Zinbarg RE. in press. Self-verification and depression in abused women. Depress Anxiety.
2. Swann Jr WB. Self-verification: bringing social reality into harmony with the self. In: Suls J, Greenwald AG. editors. Psychological perspectives on the self. Vol. II. Hillsdale, New Jersey: Erlbaum; 1983:p 33–66.
3. Swann Jr WB. Identity negotiation: where two roads meet. J Pers Soc Psychol 1987;53:1038–1051.
4. Swann Jr WB, Rentfrow PJ, Guinn J. Self-verification: the search for coherence. In: Leary M, Tangney J. Handbook of self and identity. New York: Guilford; 2003.
5. Swann Jr WB. The trouble with change: Self-verification and allegiance to the self. Psychol Sci 1997;8:177–180.
6. Lecky P. Self-consistency: a theory of personality. New York: Island Press; 1945.
7. Secord PE, Backman CW. An interpersonal approach to personality. In: Maher B. editor. Progress in experimental personality research. Vol. 2. New York: Academic Press; 1965: p 91–125.
8. Baumeister RF. Self-esteem: the puzzle of low self-regard. New York: Plenum Press; 1993.
9. Swann Jr WB, Read SJ. Acquiring self-knowledge: the search for feedback that fits. J Pers Soc Psychol 1981b;41:1119–1128.
10. Swann Jr WB, Pelham BW, Krull DS. Agreeable fancy or disagreeable truth? How people reconcile their self-enhancement and self-verification needs. J Pers Soc Psychol 1989;57:782–791.
11. Swann Jr WB, Stein-Seroussi A, Giesler B. Why people self-verify. J Pers Soc Psychol 1992b;62:392–401.
12. Swann Jr WB, Read SJ. Self-verification processes: how we sustain our self-conceptions. J Exp Soc Psychol 1981a;17: 351–372.
13. De La Ronde C, Swann Jr WB. Partner verification: restoring shattered images of our intimates. J Pers Soc Psychol 1998;75: 374–382.
14. Swann Jr WB, Hixon G, Stein-Seroussi A, Gilbert DT. The fleeting gleam of praise: behavioral reactions to self-relevant feedback. J Pers Soc Psychol 1990;59:17–26.
15. Bernichon T, Cook KE, Brown JD. Seeking self-evaluative feedback: the interactive role of global self-esteem and specific self-views. J Pers Soc Psychol 2003;84:194–204.
16. Swann Jr WB, Wenzlaff RM, Krull DS, Pelham BW. The allure of negative feedback: self-verification strivings among depressed persons. J Abnorm Psychol 1992c;101:293–306.
17. Swann Jr WB, Wenzlaff RM, Tafarodi RW. Depression and the search for negative evaluations: more evidence of the role of self-verification strivings. J Abnorm Psychol 1992d;101: 314–317.
18. Giesler RB, Josephs RA, Swann Jr WB. Self-verification in clinical depression. J Abnorm Psychol 1996;105:358–368.
19. Sedikides C, Gregg AP. Portraits of the self. In: Hogg MA, Cooper J. editors. Sage handbook of social psychology London: Sage Publications; 2003:p 110–138.
20. Sedikides C, Gregg AP. Self-enhancement: food for thought. Perspect Psychol Sci 2008;3:102–116.
21. Swann Jr WB, Hixon JG, De La Ronde C. Embracing the bitter truth: negative self-concepts and marital commitment. Psychol Sci 1992a;3:118–121.
22. Hixon JG, Swann Jr WB. When does introspection bear fruit? Self-reflection, self-insight, and interpersonal choices. J Pers Soc Psychol 1993;64:35–43.
23. Swann Jr WB, Pelham BW. Who wants out when the going gets good? Psychological investment and preference for self-verifying college roommates. J Self Identity 2002;1:219–233.
24. Swann Jr WB, De La Ronde C, Hixon JG. Authenticity and positivity strivings in marriage and courtship. J Pers Soc Psychol 1994;66:857–869.

25. Joiner TE. The price of soliciting and receiving negative feedback: self-verification theory as a vulnerability to depression. J Abnorm Psychol 1995;104:364–372.
26. Popper K. The logic of scientific discovery. Hutchinson: London; 1959.
27. Batson CD. Addressing the altruism question experimentally. In: Post SG, Underwood LG, Schloss JP, Hurlbut WB. editors. Altruism and altruistic love: science, philosophy, and religion in dialogue. New York: Oxford University Press; 2002: p 89–105.
28. Menninger K. Man against himself. New York: Harcourt, Brace, & World; 1938/1966.
29. Baumeister RF, Scher SJ. Self-defeating behavior patterns among normal individuals: review and analysis of common self-destructive tendencies. Psychol Bull 1988;104:3–22.
30. Alloy LB, Lipman AJ. Depression and selection of positive and negative social feedback: motivated preference or cognitive balance? J Abnorm Psychol 1992;101:301–313.
31. Dunning D. Self-insight: roadblocks and detours on the path to knowing thyself. New York: Psychology Press; 2005.
32. Searle J. Rationality in action. Cambridge: The MIT Press; 2001.
33. Alicke MD, Govorun O. The better-than-average effect. In: Alicke MD, Dunning DA, Krueger JI. editors. The self in social judgment. Philadelphia: Psychology Press; 2005:p 85–106.
34. Moore DA. Not so above average after all: when people believe they are worse than average and its implications for theories of bias in social comparison. Organ Behav Hum Decis Process 2007;102:42–58.
35. Dunning D, Meyerowitz JA, Holzberg AD. Ambiguity and self-evaluation: the role of idiosyncratic trait definitions in self-serving assessments of ability. J Pers Soc Psychol 1989;57: 1082–1090.
36. Seta JJ, Donaldson S, Seta CE. Self-relevance as a moderator of self-enhancement and self-verification. J Res Pers 1999;33: 442–462.
37. Swann Jr WB, Griffin JJ, Predmore S, Gaines B. The cognitive-affective crossfire: when self-consistency confronts self-enhancement. J Pers Soc Psychol 1987;52:881–889.
38. Swann Jr WB, Schroeder DG. The search for beauty and truth: a framework for understanding reactions to evaluations. Pers Soc Psychol Bull 1995;21:1307–1318.
39. Bosson J, Swann Jr WB. Self-liking, self-competence, and the quest for self-verification. Pers Soc Psychol Bull 1999;25: 1230–1241.
40. Jones SC. Self and interpersonal evaluations: Esteem theories versus consistency theories. Psychological bulletin 1973;79: 185–199.
41. Pinel EC, Swann Jr WB. The Cognitive-affective cross fire revised: affective reactions to self-discrepant evaluations. Unpublished manuscript. University of Texas at Austin; 1999.
42. Strack F, Deutsch R. Reflective and impulsive determinants of social behavior. Pers Soc Psychol Rev 2004;8:220–247.
43. Paulhus DL, Graf P, van Selst M. Attentional load increases the positivity of self-presentations. Soc Cogn 1989;7:389–400.
44. Paulhus DL, Levitt K. Desirable responding triggered by affect: automatic egotism? J Pers Soc Psychol 1987;52: 245–259.
45. Dauenheimer D, Stahlberg D, Petersen L-E. Self-discrepancy and elaboration of a self-conception as factors influencing reactions to feedback. Eur J Soc Psychol 1999;29:725–739.
46. Stahlberg D, Petersen L-E, Dauenheimer D. Preferences for and evaluation of self-relevant information depending on the elaboration of the self-schemata involved. Eur J Soc Psychol 1999;29:489–502.
47. Aronson E, Mills J. The effect of severity of initiation on liking for a group. J Abnorm Soc Psychol 1959;59:177–181.
48. Gregg AP, De Waal-Andrews W. Choices for, and perceptions of, global and specific hypothetical feedback of differential valence. Unpublished raw data, School of Psychology, University of Southampton, UK; 2007.
49. Rosenberg M. Society and the adolescent self-image. Princeton, NJ: Princeton University Press; 1965.
50. Fergusson DM, Horwood LJ, Ridder EM. Partner violence and mental health outcomes in a New Zealand birth cohort. J Marriage Fam 2005;67:1103–1119.
51. Axsom D, Cooper J. Cognitive dissonance and psychotherapy: the role of effort justification in inducing weight loss. J Exp Soc Psychol 1985;21:149–160.
52. Baumeister RF, Leary MR. The need to belong: desire for interpersonal attachments as a fundamental human motivation. Psychol Bull 1995;117:497–529.
53. Gregg AP, Hart CM, Sedikides C, Kumashiro M. Lay conceptions of modesty: a prototype analysis. Personality and Social Psychology Bulletin 2008;34:978–992.
54. Freud S. Beyond the pleasure principle. In: Strachey J. (editor). The standard edition of the complete psychological works of Sigmund Freud. Vol. 18. London: Hogarth Press; 1953:p 7–64. (Original work published 1920)