Social Psychological and Personality Science
2026, Vol. 17(3) 400–411

DOI: 10.1177/19485506251345925
journals.sagepub.com/home/spp


The Ego's Bodyguard: The Role of Personality in Self-Protective Reactions

Christoph Heine[1], Stefan C. Schmukle[2], and Michael Dufner[1]

Abstract

When faced with self-threat, people often engage in self-protective reactions. Yet not everyone does. The extent to which self-protection occurs, is thought to depend on people's personality. We put this claim to test by considering multiple personality traits from different content domains. In an experiment, participants (N = 1744) performed a bogus performance test and received negative versus positive feedback. As an indicator of self-protective reactions, we assessed participants' tendency to question the validity of the test. We found that high self-esteem and high narcissism predicted stronger self-protective reactions, particularly when these traits were assessed in the content domain targeted by the negative feedback. Contrary to our hypotheses, the self-insight motive and mindfulness also predicted stronger (and not weaker) self-protective reactions. These findings provide better understanding of the role that personality plays in motivated reasoning.

Keywords

self-protection, self-threat, personality, experiment, preregistered

Imagine failing at an exam, being rejected by a potential partner, or being criticized by your boss. Such negative feedback feels threatening, is part of everyday life, and can entail major consequences such as reducing self-esteem (Alicke & Sedikides, 2011). A self-threat occurs when negative information challenges, or contradicts a valued self-conception (Campbell & Sedikides, 1999). In classic experiments, participants took a bogus ability test and received randomly assigned negative or positive performance feedback (e.g., Pyszczynski et al., 1985). But how do people deal with self-threats?

Here, the self-protection motive—the desire to shield oneself from threat—comes into play. The motive prompts affective, cognitive, and behavioral strivings that counteract the self-threat to protect and maintain a positive self-concept. Past research has revealed its relevance and various form relevance across many contexts (Alicke & Sedikides, 2009). A typical manifestation is to question the validity of the test that underlies the negative feedback (Pyszczynski et al., 1985). However, people also evaluate the feedback provider as incompetent (Greenwald, 2002; Smalley & Stake, 1996), interpret ambiguous feedback flattering (Taylor & Crocker, 1981), take credit for successes but blame failure on external factors (Campbell & Sedikides, 1999), and deem traits more important after positive but less so after negative feedback (Tesser & Paulhus, 1983, for a review, see Sedikides, 2012).

However, what role personality plays in this context remains unclear. To address this, we conducted a comprehensive, preregistered and highly powered study, using a piloted experimental paradigm, and taking multiple personality traits from different content domains into account.

Personality Traits Relevant in Self-Protective Reactions

Because personality traits reflect characteristic ways of thinking, feeling, and behaving (Johnson, 1997; Roberts, 2009), they should influence how people generally perceive and deal with self-threat. Consistent with this idea, research found preliminary evidence that self-esteem (e.g., Beauregard & Dunning, 1998), and narcissism (e.g., Smalley & Stake, 1996) strengthen self-protective reactions, while other research speculated that personality aspects such as the self-insight motive (SIM; Heine et al., 2024), or mindfulness (Carlson, 2013) should weaken them.

Self-esteem and narcissism, both characterized by a strong self-enhancement motive (i.e., desire for positive

[1]Witten/Herdecke University, Germany
[2]Leipzig University, Germany

Corresponding Author:
Christoph Heine, Institute of Psychology and Psychotherapy, Witten/Herdecke University, Alfred-Herrhausen-Straße 50, 58448 Witten, Germany.
Email: christoph.heine@uni-wh.de

self-views; Alicke & Sedikides, 2009), should strengthen self-protective reactions for two general reasons (Raskin et al., 1991). First, they should increase the experienced threat of negative feedback. Self-esteem represents overall self-concept positivity (Rosenberg et al., 1995), while narcissism involves a grandiose self-concept (Back et al., 2013). If the self-concept is negative, then there is no positive self-view to defend and feedback should be perceived as accurate (Swann et al., 2007). Conversely, the more positive the self-concept is, the more threatening the feedback and the stronger self-protective reactions should be (Campbell & Sedikides, 1999). The second argument is that people differ in how they typically deal with negative feedback. Those who habitually self-protect will end up with positive self-views—including high self-esteem and high narcissism (e.g., Campbell & Sedikides, 1999). Consistent with this idea, the self-regulation model of narcissism describes narcissists as people who have a broad range of cognitive-behavioral strategies to quickly and routinely ward off self-threat (Morf & Rhodewalt, 2001).

In line with these arguments, initial research indicates that individuals with high self-esteem (Beauregard & Dunning, 1998; McGregor et al., 2009) and those high in grandiose narcissism (e.g., Kernis & Sun, 1994; Morf & Rhodewalt, 1993; Rhodewalt & Morf, 1998) have a pronounced proclivity to show self-protective behavior and to question the validity of a bogus test on which they (ostensibly) performed poorly (Smalley & Stake, 1996). However, there are limitations to these earlier studies. First, they measured self-protective reactions rather indirectly (Beauregard & Dunning, 1998; McGregor et al., 2009) or lacked a control condition providing non-negative feedback (Smalley & Stake, 1996). For only then can it be examined whether the derogatory evaluation of the test occurs only after self-threat and thus specifically serves to protect the self.

Second, previous research has neglected the domain-specificity of both self-esteem (Shavelson et al., 1976) and narcissism (Gebauer et al., 2012; Grosz et al., 2022). Experiencing negative feedback should be perceived as particularly threatening if it clashes with positive self-views in the threatened domain. For example, people who have built their (narcissistic) esteem on their interpersonal skills should be highly threatened by negative feedback on this particular skill. Thus, pronounced self-protective reactions should be observed among persons with high self-esteem, respectively narcissism, in the threatened domain.

Third, the role of grandiose narcissism's major subcomponents remains unclear. The narcissistic admiration and rivalry concept (NARC, Back et al., 2013) distinguishes *admiration* (assertive self-enhancement) from *rivalry* (antagonistic self-protection). Given the self-protective nature of rivalry, it should, in theory, be related to stronger self-protective reactions. However, the more consistent predictor of self-protective attributions after negative outcomes was admiration in a recent study (Kraft et al., 2024). Thus, both sub-components may be relevant for self-protection, rivalry from a theoretical and admiration from an empirical perspective.

Whereas self-esteem and narcissism should amplify self-protection, the two remaining traits of our investigation—the SIM and mindfulness—are expected to attenuate self-protective reactions. The SIM (also known as the self-assessment motive) is one of the major self-evaluation motives that shape how people select self-relevant information and draw inferences about themselves (alongside self-enhancement/self-protection and self-verification; see Sedikides & Strube, 1997). The SIM reflects a desire for accurate self-knowledge (Heine et al., 2024). People with a strong SIM are thought to consider negative self-relevant information as valuable (Gregg et al., 2011). Thus, they should suppress their self-protective tendencies because the potentially gained self-insight outweighs the threat of the negative feedback (Sedikides & Strube, 1997).

Mindfulness is characterized by non-evaluative attention to and awareness of one's current experience (Brown & Ryan, 2003; Kabat-Zinn, 1994). Mindful people are highly aware of their thoughts, feelings, and bodily sensations (Garland et al., 2015), which can help consciously recognizing self-threaten and thus avoiding automatic self-protective reactions (Carlson, 2013). Unlike the SIM, mindfulness involves the tendency of only noticing thoughts and affective states without judging or dwelling on them (Hayes-Skelton & Graham, 2013). This allows mindful people to experience aversive thoughts and emotions following negative feedback as temporary events that require no direct response. Supporting this idea, previous research indicates that when recounting past self-threatening experiences, mindful individuals display less defensive verbal behavior than persons low in mindfulness (Lakey et al., 2008). Accordingly, mindfulness should also go along with attenuated self-protective reactions in response to negative feedback.

The Current Research

This research provides the first preregistered test of self-protection main effects and its personality moderators. We hypothesized that bogus negative feedback would trigger self-protective reactions (basic hypothesis). With regard to personality moderators, we hypothesized that high self-esteem (H1a), and self-esteem in the threatened domain (H1b) would be related to stronger self-protective reactions. Grandiose narcissism (H2a), grandiose narcissism in the threatened domain (H2b), and narcissistic admiration (H2c) and rivalry (H2d) were also thought to be related to stronger self-protective reactions. Furthermore, we hypothesized that the SIM (H4) and mindfulness (H5) would be related to weaker self-protective reactions.

In a pilot study, we established an experimental paradigm that triggers self-protective reactions. Participants completed a test introduced as a measure of social sensitivity and received bogus negative or positive feedback on

their performance. A critical decision was whether negative feedback should be contrasted with positive or no feedback in a control condition. Not providing feedback in the control condition would run counter the aim of experimental studies to vary the manipulation as specifically as possible (Chester & Lasko, 2021). Without feedback, it would be unclear whether effects are driven by receiving negative feedback or by receiving feedback per se. Positive feedback in the control condition was the better control, as in this case, only the feedback negativity was varying and could thus drive the effect. Furthermore, given that positive self-estimates are common (Alicke & Govorun, 2005), the positive feedback largely aligned with people's expectation. The results of the pilot study indicated that this manipulation was successful. Those receiving negative feedback perceived the test as less valid than those receiving positive feedback (for details see the OSF).

In the main study, we extended the findings in four ways. First, we tested whether negative feedback causes a decrease in perceived test validity, compared with initial perceptions after participants had only seen a sample item of the test. Second, we considered additional manifestations of self-protection by allowing participants to (a) devalue the competence of the researchers who conduct the study and (b) to devalue the general relevance of the trait social sensitivity. Third, we asked participants to evaluate unrelated things (object ratings) to address the possibility that negative feedback leads not only to self-protective reactions but to general negativity.

Fourth, and most importantly, we investigated the personality moderators. To do so, we assessed each of the described traits and tested whether they moderate the self-protection main effect we found in the pilot study. We assessed self-esteem in a standard, general fashion, and in a domain-specific way in relation to the threatened domain social sensitivity. For the domain-specific assessment of narcissism, we relied on communal narcissism (Gebauer et al., 2012), which seems appropriate given that social sensitivity represents a trait from the communal care domain. Although this operationalization was slightly broader than a specific assessment of narcissism with regard to social sensitivity, it had the advantage that a well-validated communal narcissism measure exists.

Method

We conducted the main study following in-principle acceptance (https://osf.io/xt3q4). The approved preregistration protocol, pilot study, codebook, supplementary online materials (SOM), data, and analysis code are available on the OSF (https://osf.io/s4af3/).

Sample

We conducted a simulation study to determine the required sample size, as recommended in the structure equation modeling (SEM) framework (Muthén & Muthén, 2002). The simulation suggested that a sample size of $N = 800$ ($n = 400$ participants per group) would be needed to detect differences in the SEM's standardized regressions coefficient of $\Delta\beta = .27$ between the two experimental conditions with an alpha level of .05 and a power of 80% (for details see SOM Section 1).

The final sample included $N = 1744$ participants (502 males, 1226 females, 16 neither female nor male) aged 18 to 89 years ($M = 47.91$, $SD = 14.80$). Of these, 821 received negative and 923 received positive feedback (see SOM Section 2 for exclusions). The sample was highly educated as 2% had no or a minimal formal degree (*Hauptschulabschluss*), 16% had a medium-level degree (*Realschulabschluss*), 38% had the highest school degree (*Abitur*), and 54% held a bachelor's or master's degree.

Procedure

Data were collected online. The study design and procedure were approved by the local ethics board. Participants had to be at least 18 years old and fluent in German. They were recruited from the German online panel PsyWeb (https://psyweb.uni-muenster.de). The incentive for participation in PsyWeb studies is individualized psychological feedback.

We implemented the same paradigm and cover story as in the pilot study. The study was announced as an investigation on personality and social perception. Participants were informed that they work on a social sensitivity test that is still being validated and that should be evaluated after completion. After registration, participants completed a pretest questionnaire that included assessments of personality traits and, after seeing a sample item of the test, an assessment of initial perceived test validity.

Subsequently, participants completed 20 items from the Reading the Mind in the Eyes Test (eyes test, Baron-Cohen et al., 2001) as the announced social sensitivity test (for further details see the pilot study on the OSF). Participants were then randomly assigned into two feedback conditions. In a *positive feedback condition*, they were told that they outperformed 80 percent of a validation sample, which means that their social sensitivity is high. In a *negative feedback condition*, they were told that they were only better than 20 percent of the validation sample, which means that their social sensitivity is low (the feedback, which included a visualization on a normal distribution curve, can be found in the codebook on the OSF). After the test, participants completed a questionnaire that included the self-protection measures and the object ratings.

At the end, they received individual feedback on their perceived test validity, compared with the distributions in our pilot study. To safeguard that everyone, including those who aborted the study, knew the feedback was bogus we sent an email containing this information to all participants.

Measures

Personality Measures

Self-Esteem. We assessed self-esteem with the German version (von Collani & Herzberg, 2003) of the Rosenberg Self-Esteem Scale (Rosenberg, 1965). The scale consists of ten items (e.g., "On the whole, I am satisfied with myself") that were answered on a scale from 1 (*strongly disagree*) to 4 (*strongly agree*).

Self-Esteem in the Threatened Domain. To assess domain-specific self-esteem with regard to social sensitivity, we adapted the self-estimated social sensitivity measure from the pilot study. Participants used a range-slider to estimate their percentile in the population, based on a normal distribution displayed above the slider (similar to the later provided feedback; see the codebook on the OSF for a visualization).

As the other traits were assessed using personality scales, we also included an explorative scale-based assessment of domain-specific self-esteem (for details see preregistration protocol on the OSF). The results of this measure are summarized in the SOM Section 8 and aligned with the results of the range-slider measure.

Grandiose Narcissism. We used the short form (Ames et al., 2006) of the German (Schütz et al., 2004) Narcissistic Personality Inventory (NPI-16; Raskin & Terry, 1988), the most popular unidimensional measure of grandiose narcissism. It consists of 16 items, each presenting a pair of sentences with a high- and a low-narcissistic choice.

Grandiose Narcissism in the Threatened Domain. As an indicator of domain-specific grandiose narcissism, we measured communal narcissism using the Communal Narcissism Inventory (CNI, Gebauer et al., 2012). The scale is unidimensional and consists of 16 items (e.g., "I am the most helpful person I know") that were answered on a scale from 1 (*strongly disagree*) to 7 (*strongly agree*).

Narcissistic Admiration and Rivalry. To assess both major subcomponents of grandiose narcissism, we used the Narcissistic Admiration and Rivalry Questionnaire (Back et al., 2013). The scale comprises the factors *admiration* (e.g., "I am great") and *rivalry* (e.g., "I often get annoyed when I am criticized") and consists of 18 items that were answered on a scale from 1 (*strongly disagree*) to 6 (*strongly agree*).

Self-Insight Motive. We used the Self-Insight Motive Scale (Heine et al., 2024) to assess the SIM. The unidimensional scale consists of five items (e.g., "I want to know exactly what my strengths and weaknesses are") that were answered on a scale from 1 (*strongly disagree*) to 6 (*strongly agree*).

Mindfulness. Altgassen et al. (2023) used ant-colony optimization to select the best combination of items from the eight most commonly used mindfulness scales. We used nine items from their selected item combinations (e.g., "I try to notice my thoughts without judging them"), so that mindfulness can be assessed with high reliability in a well-fitting unidimensional model (for item selection details, see the preregistration protocol on the OSF). The items were answered on a scale from 1 (*strongly disagree*) to 7 (*strongly agree*).

Outcome Measures

Perceived Test Validity. As the main self-protection measure, participants evaluated the suitability of the eyes test to measure social sensitivity on the six items we had selected in the pilot study (e.g., "The test is suitable for the assessment of social sensitivity," for details see the pilot study on the OSF). The items were answered on a scale from 1 (*strongly disagree*) to 5 (*strongly agree*).

Additional Outcome Measures. As a second self-protection measure, participants evaluated the *competence of the researcher* on six items (e.g., "I have no confidence in the expertise of the researchers"), of which we finally selected four items (see SOM Section 4). As a third self-protection measure, we included six items that asked participants to rate the *importance of social sensitivity* (e.g., "Social sensitivity is a very important skill"). As a final measure that cannot be utilized for self-protective purposes, we used an object rating task (Rau et al., 2020), which allowed us to test for a general negativity bias following negative feedback. In that task, participants indicated how much they liked photographs of five nonsocial objects (e.g., a mug and a hanger) with three items each (e.g., "I like this object"). All items of the additional outcome measures can be found in the codebook on the OSF and were answered on a scale from 1 (*strongly disagree*) to 5 (*strongly agree*).

Statistical Analysis

Analyses were conducted using the *lavaan* package (Rosseel, 2012) in R (R Core Team, 2020). For the basic hypothesis, we used likelihood ratio tests to examine whether latent means of self-protection outcomes and object ratings differed between feedback conditions. In addition, we estimated a latent change model to test whether perceived test validity specifically decreased after negative feedback compared with the initial evaluation (McArdle, 2009; see SOM Section 7 for a visualization).

For the moderation hypotheses, we included personality traits as predictors of the outcomes in SEMs (Figure 1 illustrates this for H1a). Predictions after positive feedback provided a "baseline" for trait-outcome relations unaffected by self-threat, against which predictions after negative feedback were compared. As lower perceived test validities

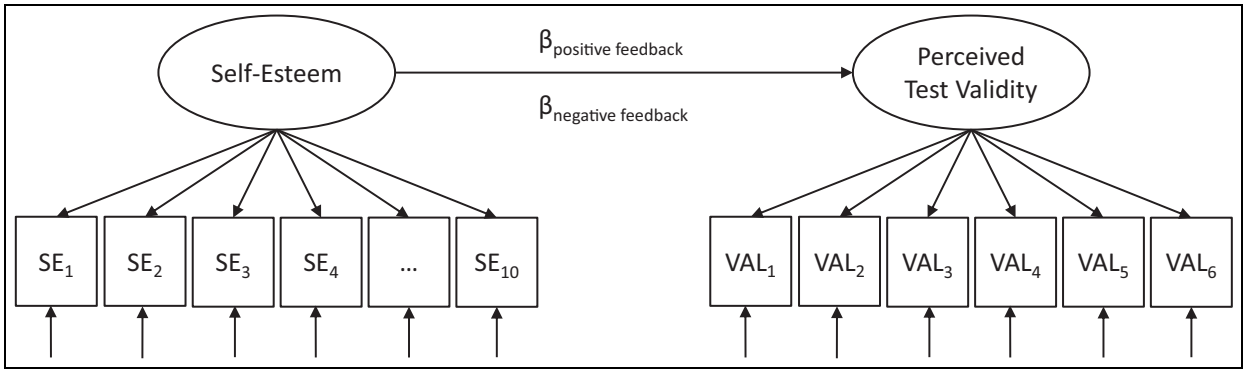

Figure 1. *Schematic Structure Equation Model for the Test of Hypothesis 1a*
Note. SE_1 to SE_{10}: Items from the Rosenberg Self-Esteem Scale, VAL_1 to VAL_6: Items from the perceived test validity, the model was estimated for a multigroup setting with an invariant measurement model of the respective trait and the perceived test validity between positive and negative feedback. A significant difference in the predictor effect (β) between positive and negative feedback indicates a relation of the trait to self-protective reactions.

represented stronger self-protective reactions, more negative predictions of the perceived test validity after negative compared with positive feedback indicated that the trait was related to stronger self-protective reactions (as suggested for self-esteem and narcissism).

We estimated separate SEMs for each for each personality trait and used a likelihood ratio test to compare two models with freely estimated versus equality-constrained regression coefficients across feedback conditions. A significant test indicated that a trait's prediction was more negative (or positive) in the negative feedback condition, suggesting its relation to self-protective reactions. To account for testing the same effect on three self-protection outcomes, we applied a false discovery rate correction (Benjamini & Hochberg, 1995). We finally combined all traits that moderated self-protective reactions in a single model to assess their incremental contributions (preconditions for these combined models, including trait correlations and variance inflation factors, are provided in the SOM Sections 2 and 5).

Results

Preliminary analyses are available in the SOM, including measurement models, reliabilities, descriptive statistics, and correlations of the personality traits (Section 2), and measurement models, invariance tests across feedback conditions, and reliabilities of the outcomes (Section 3). Results for the basic hypothesis are summarized in Figure 2.

Consistent with the pilot study, participants who received negative feedback perceived the eyes test as less valid than those who received positive feedback ($d = -.84$, $p < .001$). We also found a significant difference in latent changes between conditions compared with initial estimates, $\chi^2(1) = 363.678$, $p < .001$. Perceived test validity slightly increased in the positive condition ($d = .12$, $p = .003$), but substantially decreased in the negative condition ($d = -.89$, $p < .001$), indicating a self-protective reaction after negative feedback.

Similarly, participants who received negative feedback evaluated the researchers' competence lower ($d = -.62$, $p < .001$) and social sensitivity as less important compared with those who received positive feedback ($d = -.19$, $p < .001$). The effect for the importance of social sensitivity was smaller, possibly because lower importance ratings might cause cognitive dissonance (Festinger & Carlsmith, 1959) among PsyWeb participants who voluntarily invested time to learn about social sensitivity. As a result, they may have favored alternative strategies, such as questioning the validity of the test or the researchers' competence. Finally, participants who received negative feedback also rated the objects more negatively ($d = -.25$, $p < .001$), suggesting that negative feedback did not only cause self-protective reactions, but also more negative evaluations in general. Yet, this general negativity effect could not fully explain reactions to self-threatening feedback, as this effect was a lot smaller than those for perceived test validity and researcher competence evaluations.

We then tested the moderating effects of personality, with results summarized in Table 1. Our first hypothesis was supported as global self-esteem (H1a) moderated group differences in perceived test validity and the importance of social sensitivity. Domain-specific self-esteem, measured via self-estimated social sensitivity (H1b), moderated all three self-protection outcomes, with a noteworthy effect on the main outcome perceived test validity ($\Delta\beta = -.56$, $p < .001$). In all cases, the moderation aligned with the expectations, showing more negative predictor effects of self-esteem after negative compared with positive feedback.

Our second hypothesis was also largely supported. Grandiose narcissism moderated all three self-protection

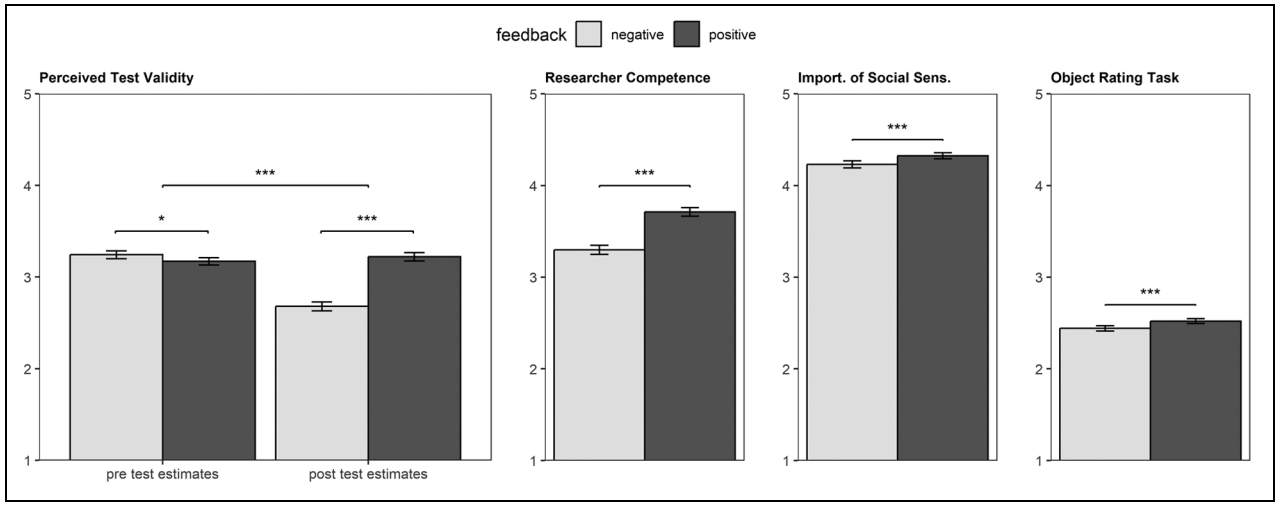

Figure 2. *Results of the Basic Hypothesis Displayed by Outcome*
Note. Import. of Social Sens. = Importance of social sensitivity.

outcomes in the expected direction, both in its global (H2a) and domain-specific (H2b) forms. Again, we found a particularly strong moderator effect of domain-specific narcissism on the main outcome perceived test validity ($\Delta\beta = -.45$, $p < .001$). Distinguishing admiration (H2c) from rivalry (H2d), only admiration moderated all outcomes whereas rivalry moderated none of them. This result indicates that admiration, and not rivalry, drives narcissists' self-protection tendencies.

The self-insight motive (H3) and mindfulness (H4) also moderated outcomes, but in the opposite direction from what was expected in the hypotheses. As for self-esteem and narcissism, individuals with a strong SIM evaluated the test as less valid and researchers as less competent when receiving negative feedback, compared with those with a weaker SIM and compared with positive feedback. Likewise, individuals with high mindfulness also had a stronger tendency to evaluate the test as less valid and social sensitivity as less important after negative feedback. The results thus suggest these traits may enhance, rather than attenuate, self-protective reactions, contradicting our original hypotheses.

As we found a group difference for the object ratings (see above), we looked at whether personality also moderated this effect. However, no moderations emerged, indicating that personality specifically moderated self-protective reactions—and not general negativity.

Finally, we estimated a combined model for each outcome, including all personality traits for which we found significant moderator effects. Domain-specific self-esteem and domain-specific narcissism remained as key moderators of perceived test validity (see Table 1, visualized in Figure 3). For researcher competence evaluations, domain-specific self-esteem and the SIM remained as significant moderators. Domain-specific self-esteem and globally assessed grandiose narcissism remained as moderators of the importance of social sensitivity. These results highlight the critical role of self-esteem and narcissism for self-protection, particularly in their domain-specific expressions.

We also conducted three preregistered supplementary analyses. First, we tested our hypotheses with classic linear models using mean scores of the considered variables, which produced comparable results to the original analyses (see SOM Section 4). Second, we re-ran our original analyses while controlling for covariates (performance on the eyes test, certainty of the social-sensitivity self-estimate, and prior test knowledge) and found that this did not alter the results to a substantial degree (see SOM Section 5). Third and finally, we included the personality traits as predictors in the latent change score model (see above) and found the same moderator effects for latent changes in perceived test validity (see SOM Section 6).

Discussion

The current research was the first preregistered, highly powered test of the hypothesis that negative feedback elicits self-protective reactions. We incorporated three outcomes—perceived test validity, researcher competence evaluations, and importance ratings of social sensitivity—and consistently found support for this hypothesis, underscoring that self-protective reactions are a real and relevant phenomenon (Nosek & Lakens, 2014). A practical implication for self-protection research is that the design, procedure, and outcome measures, which demonstrated invariance across feedback conditions, could serve as a standard experimental paradigm for future studies. Establishing standardized paradigms is central for a mature cumulative science (Allen et al., 2017).

Table 1. *Personality Moderator Effects on the Group Differences Between in the Self-Protection Outcomes and Object Ratings*

	Perceived test validity			Researcher competence evaluation			Importance of social sensitivity			Object rating task		
Moderator	β_{pos}	β_{neg}	χ^2_{single} ($\chi^2_{comb.}$)	β_{pos}	β_{neg}	χ^2_{single} ($\chi^2_{comb.}$)	β_{pos}	β_{neg}	χ^2_{single} ($\chi^2_{comb.}$)	β_{pos}	β_{neg}	χ^2_{single} ($\chi^2_{comb.}$)
H1a Self-esteem	.03	−.09*	5.22* (0.88)	.01	.00	0.01 (-)	.31***	.12**	10.07** (0.39)	.05	.04	0.06 (-)
H1b Self-esteem in the threatened domain	.19***	−.38***	131.37*** (15.71***)	.12***	−.17***	32.10*** (13.40***)	.20***	.07	5.80** (13.24***)	.06	.03	0.21 (-)
H2a Grandiose narcissism	.12**	−.08	9.69** (1.21)	.09*	−.12**	11.90*** (1.88)	.20***	.01	7.88** (3.32)	.01	−.14**	4.45 (-)
H2b Grandiose narcissism in the threatened domain	.25***	−.20***	68.51*** (9.63**)	.14***	.00	6.39** (.01)	.23***	.09*	5.37* (3.10)	.21***	.17***	0.49 (-)
H2c Narcissistic admiration	.16***	−.06	16.20*** (0.08)	.09*	−.06	7.01** (.60)	.19***	.03	6.95** (3.97*)	.12*	.01	2.40 (-)
H2d Narcissistic rivalry	.10*	.08*	0.23 (-)	−.02	−.01	0.03 (-)	−.18***	−.14**	0.18 (-)	−.05	−.07*	0.15 (-)
H3 Self-Insight Motive	.19***	.01	10.65*** (0.21)	.19***	.01	10.88*** (7.84**)	.27***	.24***	0.08 (-)	.08	.13	0.59 (-)
H4 Mindfulness	.03	−.11*	5.92* (.004)	.01	−.04	0.80 (-)	.25***	.11*	5.35* (0.59)	.18***	.16**	0.08 (-)

Note. βpos/βneg = predictor effect in the positive/negative feedback condition; χ^2 single (χ^2 comb.) = likelihood ratio test for a difference in the personality predictor effects between feedback conditions testing the moderation hypotheses; values in brackets represent tests from the combined model that simultaneously included all personality moderators that were significant in the single comparisons; *p*-values of the likelihood ratio test are adjusted for multiple testing.
*$p < .05$. **$p < .01$. ***$p < .001$.

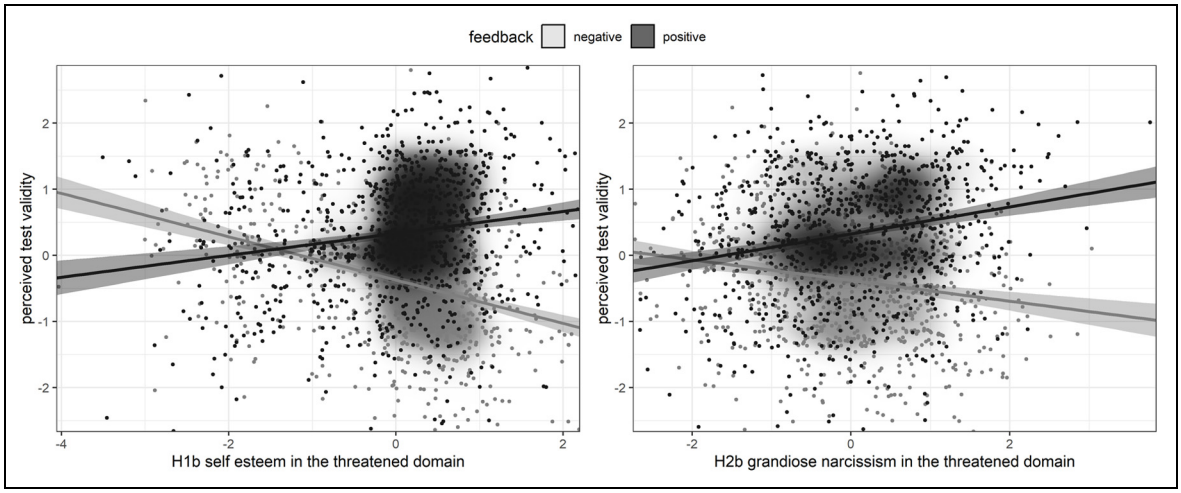

Figure 3. *Central Moderators of the Perceived Test Validity*
Note. Participants' individual data points are slightly moved in random directions to reduce their overlap; areas with the highest densities in both conditions are slightly highlighted.

Most importantly, the results for the first time contributed robust evidence linking personality traits to self-protection. Regarding self-esteem, our results support the accumulated preliminary evidence (e.g., Beauregard & Dunning, 1998). We expanded the theoretical basis by differentiating between overall self-concept positivity and domain-specific self-perceptions and found that people with a positive self-concept in the particular threatened domain are prone to self-protective reactions. The findings suggest a potential downside to high self-esteem (which is typically considered adaptive, Baumeister et al., 2003), as individuals may resist negative information that challenge their self-views. Given that people invest in domains where they feel competent and even base career paths on these perceived strengths, acknowledging the implications of negative feedback can be highly crucial. Our results highlight that balancing self-esteem maintenance with incorporating negative feedback likely is a challenging task for many people (Crocker & Park, 2004).

The results also support the often-suggested link of grandiose narcissism to self-protection. A central claim of the NARC is that rivalry drives narcissists' self-protective reactions (Back et al., 2013). Yet, our direct empirical test indicates that this might not be the case, which parallels recent findings (Kraft et al., 2024). Only the admiration component consistently predicted stronger self-protection, suggesting that rivalry may be better understood as a general antagonistic tendency (Moshagen et al., 2018)—one that is less directly tied to self-threat responses. In contrast, the domain-specific narcissism measure demonstrated a strong moderator effect, highlighting that the content-wise differentiation of the narcissism construct is a fruitful approach (Gebauer et al., 2012; Grosz et al., 2022).

For the remaining two traits, the SIM and mindfulness, we initially expected an attenuation of self-protective reactions. Individuals with a strong SIM are believed to place high value on negative self-relevant information, which should, in theory, reduce their tendency toward self-protection. Contrary to these expectations, however, individuals with a strong SIM showed more self-protection. This finding may explain why individuals with a strong SIM do not have more accurate self-perceptions compared with those with a weaker SIM (Heine et al., 2024). Even those motivated by self-insight may struggle to achieve greater self-knowledge because, like others—and potentially even more so—they engage in self-protection when confronted with self-threats.

Similarly, we anticipated that mindfulness would reduce self-protection, as mindfulness is thought to help people perceive negative feedback as a transient event that does not require immediate defense. However, our results again contradicted these expectations, as mindful individuals showed stronger self-protective reactions. A likely explanation for this lies in the conceptualization of mindfulness as an ability (Bishop et al., 2004), which is assessed through self-reports (e.g., Baer et al., 2004). Yet, self-reports are often not valid measures of abilities (Zell & Krizan, 2014) and may instead reflect a form of communal self-enhancement in the context of mindfulness (Gebauer et al., 2018). The present results on communal narcissism indicate that communal self-enhancement fosters self-protective reactions in response to negative communal feedback. As mindfulness was positively related to communal narcissism, this likely explains the relationship between mindfulness and self-protection. These findings add to recent critiques regarding the conceptualization and measurement of mindfulness (Altgassen et al., 2023). Future research could aim

to develop valid mindfulness measures that are not based on self-reports.

A strength of the current research is that it considered all traits simultaneously that should in theory be linked to self-protective reactions. The strongest and most robust effects emerged for the domain-specific assessments of self-esteem and narcissism, suggesting that these traits warrant further in-depth investigation into the underlying mechanisms of self-protection and their relationship to personality.

This research also has practical implications. Self-protective reactions, while sometimes helpful for maintaining a positive self-image (Taylor & Brown, 1988), can lead individuals to reject highly self-relevant information, such as feedback on maladaptive behaviors. In clinical settings, such defensive tendencies may hinder the therapeutic goal of achieving clarity (Grawe, 2004), especially in domains tied to strongly positive self-views. This may shed light on why some individuals struggle to gain insights necessary for improving personal well-being or making therapeutic progress. Addressing these tendencies could improve the effectiveness of interventions, and outcomes for individuals resistant to self-reflection or behavioral change.

Limitations and Future Directions

In the present research, we focused on personality's role in self-protective reactions using a classic face-valid paradigm that has been applied in similar versions. Still, other self-evaluation motives—such as self-verification (Swann & Read, 1981)—may have influenced perceived test validity. For example, people with low self-esteem might have perceived the test as particularly valid after receiving negative feedback because that feedback aligned with their negative self-concept. Even though the present findings are most plausible driven by self-protection processes, future work should clarify the exact ongoing psychological process.

Another limitation is that we only provided social sensitivity feedback, which means that the results do not provide definite evidence on whether the domain-specificity represents the underlying causal mechanism. Future studies could systematically vary the domains in which feedback is provided (e.g., social sensitivity and intelligence) and assess corresponding domain-specific self-esteem and narcissism, to determine whether they are consistently most relevant for self-protective reactions in the respective threatened domain.

Conclusion

The present research took an individual differences perspective on self-protective reactions, a research topic that has thus far been studied almost exclusively from the perspective of experimental social psychology. Building interactionist bridges to the field of personality science can stimulate interdisciplinary research on motivated reasoning and motivated social cognition. We hope that such integrative approaches will result in a better understanding of how the interplay between personal and situational factors shape behavior.

Acknowledgment

We wish to thank Anna Krebs for collecting the data of the pilot study.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The study was funded by the German Research Foundation (DFG) via Grant DU 1641/5-1.

Ethical Approval

The approved preregistration protocol (https://osf.io/xt3q4), pilot study, codebook, supplementary online materials (SOM), data, and analysis code can be found on the OSF (https://osf.io/s4af3/).

ORCID iDs

Christoph Heine https://orcid.org/0000-0001-7884-3022
Stefan C. Schmukle https://orcid.org/0000-0002-6279-9618
Michael Dufner https://orcid.org/0000-0002-8354-6193

Supplemental Material

Supplemental material for this article is available online.

References

Allen, R. (2017). *Statistics and experimental design for psychologists: A model comparison approach*. World Scientific Publishing. https://doi.org/10.1142/Q0019

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

Alicke, M. D., & Sedikides, C. (2009). Self-enhancement and self-protection: What they are and what they do. *European Review of Social Psychology*, *20*(1), 1–48. https://doi.org/10.1080/10463280802613866

Alicke, M. D., & Sedikides, C. (Eds.). (2011). *Handbook of self-enhancement and self-protection*. The Guilford Press.

Altgassen, E., Geiger, M., & Wilhelm, O. (2023). Do you mind a closer look? A jingle-jangle fallacy perspective on mindfulness. *European Journal of Personality*, *38*(2), 365–387. https://doi.org/10.1177/08902070231174575

Ames, D. R., Rose, P., & Anderson, C. P. (2006). The NPI-16 as a short measure of narcissism. *Journal of Research in Personality*, *40*(4), 440–450. https://doi.org/10.1016/j.jrp.2005.03.002

Back, M. D., Küfner, A. C. P., Dufner, M., Gerlach, T. M., Rauthmann, J. F., & Denissen, J. J. A. (2013). Narcissistic admiration and rivalry: Disentangling the bright and dark sides of narcissism. *Journal of Personality and Social Psychology*, *105*(6), 1013–1037. https://doi.org/10.1037/a0034431

Baer, R. A., Smith, G. T., & Allen, K. B. (2004). Assessment of mindfulness by self-report: The Kentucky Inventory of Mindfulness Skills. *Assessment*, *11*(3), 191–206. https://doi.org/10.1177/1073191104268029

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(2), 241–251.

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), 1–44. https://doi.org/10.1111/1529-1006.01431

Beauregard, K. S., & Dunning, D. (1998). Turning up the contrast: Self-enhancement motives prompt egocentric contrast effects in social judgments. *Journal of Personality and Social Psychology*, *74*(3), 606–621. https://doi.org/10.1037/0022-3514.74.3.606

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300.

Bishop, S. R., Lau, M., Shapiro, S., Carlson, L., Anderson, N. D., Carmody, J., Segal, Z. V., Abbey, S., Speca, M., Velting, D., & Devins, G. (2004). Mindfulness: A proposed operational definition. *Clinical Psychology*, *11*(3), 230–241. https://doi.org/10.1093/clipsy.bph077

Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology*, *84*(4), 822–848. https://doi.org/10.1037/0022-3514.84.4.822

Campbell, W. K., & Sedikides, C. (1999). Self-threat magnifies the self-serving bias: A meta-analytic integration. *Review of General Psychology*, *3*(1), 23–43. https://doi.org/10.1037/1089-2680.3.1.23

Carlson, E. N. (2013). Overcoming the barriers to self-knowledge: Mindfulness as a path to seeing yourself as you really are. *Perspectives on Psychological Science*, *8*(2), 173–186. https://doi.org/10.1177/1745691612462584

Chester, D. S., & Lasko, E. N. (2021). Construct validation of experimental manipulations in social psychology: Current practices and recommendations for the future. *Perspectives on Psychological Science*, *16*(2), 377–395. https://doi.org/10.1177/1745691620950684

Crocker, J., & Park, L. E. (2004). The costly pursuit of self-esteem. *Psychological Bulletin*, *130*(3), 392–414. https://doi.org/10.1037/0033-2909.130.3.392

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *The Journal of Abnormal and Social Psychology*, *58*(2), 203–210. https://doi.org/10.1037/h0041593

Garland, E. L., Farb, N. A. R., Goldin, P., & Fredrickson, B. L. (2015). Mindfulness broadens awareness and builds eudaimonic meaning: A process model of mindful positive emotion regulation. *Psychological Inquiry*, *26*(4), 293–314. https://doi.org/10.1080/1047840X.2015.1064294

Gebauer, J. E., Nehrlich, A. D., Stahlberg, D., Sedikides, C., Hackenschmidt, A., Schick, D., Stegmaier, C. A., Windfelder, C. C., Bruk, A., & Mander, J. (2018). Mind-body practices and the self: Yoga and meditation do not quiet the ego but instead boost self-enhancement. *Psychological Science*, *29*(8), 1299–1308. https://doi.org/10.1177/0956797618764621

Gebauer, J. E., Sedikides, C., Verplanken, B., & Maio, G. R. (2012). Communal narcissism. *Journal of Personality and Social Psychology*, *103*(5), 854–878. https://doi.org/10.1037/a0029629

Grawe, K. (2004). *Psychological therapy*. Hogrefe Publishing.

Greenwald, A. G. (2002). Constructs in student ratings of instructors. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 277–297). Lawrence Erlbaum.

Gregg, A. P., Sedikides, C., & Gebauer, J. E. (2011). Dynamics of identity: Between self enhancement and self-assessment. In S. J. Schwartz, K. Luyckx, & V. L. Vignoles (Eds.), *Handbook of identity theory and research* (pp. 305–327). Springer. https://doi.org/10.1007/978-1-4419-7988-9_14

Grosz, M. P., Hartmann, I., Dufner, M., Leckelt, M., Gerlach, T. M., Rauthmann, J. F., Denissen, J. J. A., Küfner, A. C. P., & Back, M. D. (2022). A process × domain assessment of narcissism: The domain-specific narcissistic admiration and rivalry questionnaire. *Assessment*, *29*(7), 1482–1495. https://doi.org/10.1177/10731911211020075

Hayes-Skelton, S., & Graham, J. (2013). Decentering as a common link among mindfulness, cognitive reappraisal, and social anxiety. *Behavioural and Cognitive Psychotherapy*, *41*(3), 317–328. https://doi.org/10.1017/S1352465812000902

Heine, C., Schmukle, S. C., & Dufner, M. (2024). The quest for genuine self-knowledge: An investigation into individual differences in the self-insight motive. *European Journal of Personality*. Advance online publication. https://doi.org/10.1177/08902070241272184

Johnson, J. A. (1997). Units of analysis for the description and explanation of personality. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 73–93). Academic Press. https://doi.org/10.1016/B978-012134645-4/50004-4

Kabat-Zinn, J. (1994). *Wherever you go, there you are: Mindfulness meditation in everyday life*. Hyperion Books.

Kernis, M. H., & Sun, C. R. (1994). Narcissism and reactions to interpersonal feedback. *Journal of Research in Personality*, *28*(1), 4–13. https://doi.org/10.1006/jrpe.1994.1002

Kraft, L., Zimmermann, J., Schmukle, S. C., Czarna, A. Z., Sekerdej, M., & Dufner, M. (2024). Who gets the credit for success and the blame for failure? On the links between narcissism and self- and group-serving biases. *European Journal of Personality*, *38*(3), 476–493. https://doi.org/10.1177/08902070231199366

Lakey, C. E., Kernis, M. H., Heppner, W. L., & Lance, C. E. (2008). Individual differences in authenticity and mindfulness as predictors of verbal defensiveness. *Journal of Research in Personality*, *42*(1), 230–238. https://doi.org/10.1016/j.jrp.2007.05.002

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, *60*, 577–605. https://doi.org/10.1146/annurev.psych.60.110707.163612

McGregor, I., Nash, K. A., & Inzlicht, M. (2009). Threat, high self-esteem, and reactive approach-motivation: Electroencephalographic evidence. *Journal of Experimental Social Psychology*, *45*(4), 1003–1007. https://doi.org/10.1016/j.jesp.2009.04.011

Morf, C. C., & Rhodewalt, F. (1993). Narcissism and self-evaluation maintenance: Explorations in object relations. *Personality and Social Psychology Bulletin*, *19*(6), 668–676. https://doi.org/10.1177/0146167293196001

Morf, C. C., & Rhodewalt, F. (2001). Unraveling the paradoxes of narcissism: A dynamic self-regulatory processing model. *Psychological Inquiry*, *12*(4), 177–196. https://doi.org/10.1207/S15327965PLI1204_1

Moshagen, M., Hilbig, B. E., & Zettler, I. (2018). The dark core of personality. *Psychological Review*, *125*(5), 656–688. https://doi.org/10.1037/rev0000111

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, *9*(4), 599–620. https://doi.org/10.1207/S15328007SEM0904_8

Nosek, B. A., & Lakens, D. (2014). Registered reports. *Social Psychology*, *45*(3), 137–141. https://doi.org/10.1027/1864-9335/a000192

Pyszczynski, T., Greenberg, J., & LaPrelle, J. (1985). Social comparison after success and failure: Biased search for information consistent with a self-serving conclusion. *Journal of Experimental Social Psychology*, *21*(2), 195–211. https://doi.org/10.1016/0022-1031(85)90015-0

Raskin, R., Novacek, J., & Hogan, R. (1991). Narcissism, self-esteem, and defensive self-enhancement. *Journal of Personality*, *59*(1), 19–38. https://doi.org/10.1111/j.1467-6494.1991.tb00766.x

Raskin, R., & Terry, H. (1988). A principal-components analysis of the Narcissistic Personality Inventory and further evidence of its construct validity. *Journal of Personality and Social Psychology*, *54*(5), 890–902. https://doi.org/10.1037/0022-3514.54.5.890

Rau, R., Nestler, W., Dufner, M., & Nestler, S. (2020). Seeing the best or worst in others: A measure of generalized other-perceptions. *Assessment, 28*(8), 1897–1914. https://doi.org/10.1177/1073191120905015

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rhodewalt, F., & Morf, C. C. (1998). On self-aggrandizement and anger: A temporal analysis of narcissism and affective reactions to success and failure. *Journal of Personality and Social Psychology*, *74*(3), 672–685. https://doi.org/10.1037/0022-3514.74.3.672

Roberts, B. W. (2009). Back to the future: Personality and assessment and personality development. *Journal of Research in Personality*, *43*(2), 137–145. https://doi.org/10.1016/j.jrp.2008.12.015

Rosenberg, M. (1965). Rosenberg self-esteem scale. *Journal of Religion and Health*, *59*(1), 381–398.

Rosenberg, M., Schooler, C., Schoenbach, C., & Rosenberg, F. (1995). Global self-esteem and specific self-esteem: Different concepts, different outcomes. *American Sociological Review*, *60*(1), 141–156. https://doi.org/10.2307/2096350

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36. https://doi.org/10.18637/jss.v048.i02

Schütz, A., Marcus, B., & Sellin, I. (2004). Die Messung von Narzissmus als Persönlichkeitskonstrukt: Psychometrische Eigenschaften einer Lang- und einer Kurzform des Deutschen NPI (Narcissistic Personality Inventory) [Measuring narcissism as a personality construct: Psychometric properties of a long and a short version of the German Narcissistic Personality Inventory]. *Diagnostica*, *50*(4), 202–218. https://doi.org/10.1026/0012-1924.50.4.202

Sedikides, C. (2012). Self-protection. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (pp. 327–353). The Guilford 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, pp. 209–269). Academic Press. https://doi.org/10.1016/S0065-2601(08)60018-0

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-Concept: Validation of construct interpretations. *Review of Educational Research, 46*(3), 407–441. https://doi.org/10.3102/00346543046003407

Smalley, R. L., & Stake, J. E. (1996). Evaluating sources of ego-threatening feedback: Self-esteem and narcissism effects. *Journal of Research in Personality*, *30*(4), 483–495. https://doi.org/10.1006/jrpe.1996.0035

Swann, W. B. Jr., Chang-Schneider, C., & Larsen McClarty, K. (2007). Do people's self-views matter? Self-concept and self-esteem in everyday life. *American Psychologist*, *62*(2), 84–94. https://doi.org/10.1037/0003-066X.62.2.84

Swann, W. B. Jr., & Read, S. J. (1981). Self-verification processes: How we sustain our self-conceptions. *Journal of Experimental Social Psychology*, *17*(4), 351–372. https://doi.org/10.1016/0022-1031(81)90043-3

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103*(2), 193–210. https://doi.org/10.1037/0033-2909.103.2.193

Taylor, S. E., & Crocker, J. (1981). Schematic bases of social information processing. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), *Social cognition: The Ontario symposium* (pp. 89–134). Lawrence Erlbaum.

Tesser, A., & Paulhus, D. (1983). The definition of self: Private and public self-evaluation management strategies. *Journal of Personality and Social Psychology*, *44*(4), 672–682. https://doi.org/10.1037/0022-3514.44.4.672

von Collani, G., & Herzberg, P. Y. (2003). Eine revidierte Fassung der deutschsprachigen Skala zum Selbstwetgefühl von Rosenberg [A revised version of the German adaptation of Rosenberg's Self-Esteem Scale]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, *24*(1), 3–7. https://doi.org/10.1024/0170-1789.24.1.3

Zell, E., & Krizan, Z. (2014). Do people have insight into their abilities? A metasynthesis. *Perspectives on Psychological Science*, *9*(2), 111–125. https://doi.org/10.1177/1745691613518075

Author Biographies

Christoph Heine is an academic researcher at Witten/Herdecke University. His research primarily deals with motivated self-perception and self-insight.

Stefan C. Schmukle is a full professor at the University of Leipzig. He is interested in the analysis of determinants, changes and consequences of personality in large-scale studies.

Michael Dufner is a full professor at Witten/Herdecke University. His research primarily deals with self-enhancement and motive dispositions.

Handling Editor: Alexa Tullett