The Viability and Validity of the Authentic and Hubristic Pride Scales: Reply to Dickens and Murphy (2023)

Jessica L. Tracy¹, Eric J. Mercadante¹, and Richard W. Robins²

¹ Department of Psychology, University of British Columbia
² Department of Psychology, University of California, Davis

Dickens and Murphy (2023) claim that the Authentic and Hubristic Pride (i.e., AP/HP) scales (Tracy & Robins, 2007), which we developed and validated over 15 years ago, do not validly assess the theoretical constructs of authentic and hubristic pride (e.g., Tracy & Robins, 2004a, 2007). These authors further call for the development of new measures based on a top-down approach, which would incorporate the theory into scale items. Although we appreciate Dickens and Murphy’s emphasis on the need for valid assessment tools in this important research domain, we disagree with their conclusion that the extant scales are “fundamentally invalid.” Here, we explain why a top-down approach would not be preferable to the bottom-up one we used and review the relatively large body of evidence supporting the validity of the extant AP/HP scales. Dickens and Murphy also raised several concerns regarding the HP scale specifically; most of these, as we explain, are either incorrect, exaggerated, or valid concerns but not ones that invalidate the HP scale. Nonetheless, we agree with Dickens and Murphy’s suggestion that the AP/HP scales could be improved, and we echo their call for future research in this vein. Finally, we recommend that scholars seeking to advance the field in this way adopt the “living document” approach advocated by Gerasimova (2022).

Keywords: authentic pride, hubristic pride, scales

As the authors of the first empirical and theoretical work on authentic and hubristic pride and the developers of the only extant scales available to measure these constructs (Tracy & Robins, 2004a, 2007), we appreciate the effort that Dickens and Murphy (2023) put into reviewing prior research on the two-facet model of pride. We take issue, however, with Dickens and Murphy’s conclusion that the authentic and hubristic pride (i.e., AP and HP) scales are “invalid for the purpose of investigating a two-facet model of pride.” We address Dickens and Murphy’s concerns with the scales below but, given that theory and measurement are inextricably linked, we first consider how the scales fit within the broader context of our theoretical model.

Our two-facet model of pride (Tracy & Robins, 2004a, 2007) includes four interrelated components, which are as follows:

1. The theoretical prediction and empirical finding that pride is not a unitary construct but rather involves two semantically and experientially distinct facets.

2. The conceptual interpretation of these facets, one of which represents an adaptive and the other a maladaptive form of pride (i.e., authentic and hubristic pride, respectively).

3. The overarching, multifaceted theoretical conceptualization of the two facets in terms of (a) the specific processes that elicit them (including attribution processes); (b) their complementarity with guilt and shame; (c) their connections to self-esteem, narcissism, and related self-evaluative processes; (d) their presumed cross-cultural universality; and (e) their presumed evolutionary function in terms of status, prestige, dominance, and related constructs.

4. The claim that the AP/HP scales validly assess the two facets (Tracy & Robins, 2004a, 2007) of pride.

Dickens and Murphy challenged Component 4 exclusively; they did not directly address the other components of the model except to assert that the AP/HP scales do not provide adequate tests of the theoretical ideas suggested in Component 3. Nonetheless, critiques of the theoretical model have been raised previously, and we agree with several of these. Most notably, Holbrook et al. (2014) noted that extant studies do not consistently support the suggestion that AP is elicited by effort attributions and...
that HP is elicited by ability attributions (see Tracy & Robins, 2014). Dickens and Murphy use this point to argue that the AP/HP scales are invalid, but an equally viable interpretation is that our original model was incorrect on this particular point. In this view, AP and HP are not as analogous to guilt and shame as we initially theorized (Tracy & Robins, 2004a). Although our initial hypotheses about the two facets’ distinct attributional antecedents stemmed from the speculation (first suggested by Tangney et al., 1989; see also Lewis, 2000) that a pair of positive self-conscious emotions might parallel guilt and shame, it has become clear that mapping of the parallel pairs is far from perfect (Dickens & Robins, 2022; Holbrook et al., 2014). Given data that have emerged since our original hypotheses, we acknowledge this point, as we did 8 years ago (Mercadante et al., 2021; Tracy & Robins, 2014). As we wrote 8 years ago, “We agree that there is reason to question the previously reported attribution patterns for AP and HP; this is an issue we initially raised in our article reporting these patterns (see Tracy & Robins, 2007, pp. 522–523). Indeed ... the correlational results that emerged from Holbrook and colleagues’ (2014) work and from Studies 3 and 5 in Tracy & Robins (2007) can be considered to provide, at best, weak and, at worst, inconsistent, support for our appraisal model” (p. 35).

In our work since 2014, we refer to the attributional distinction in relative terms and regarding specific results; most notably, the finding from an experimental study that AP was more likely to be elicited by effort attributions than by ability attributions, whereas HP was more likely to be elicited by ability attributions than by effort attributions (Tracy & Robins, 2007); the replication of this pattern in a sample of Chinese university students (Shi et al., 2015); and a study showing that observers use the effort/ability attribution distinction to determine whether the same pride nonverbal display conveys AP or HP (Tracy & Prehn, 2012). We also have referred readers to evidence against the attributional distinction (see, e.g., Mercadante et al., 2021; Tracy et al., in press). We no longer treat this distinction as a primary defining or distinguishing feature of the facets, but instead focus on other documented differences like their divergent external correlates with personality traits, social behaviors, and psychopathologies (e.g., Mercadante et al., 2021; Tracy et al., 2020, in press; Witkower et al., 2020).

As for the primary focus of Dickens and Murphy’s article—the validity of the AP/HP scales—we agree that more data are needed to support the scales’ validity, but we see no reason to conclude from extant data that they are invalid. Dickens and Murphy make several arguments to support this claim, and we take issue with each. First, Dickens and Murphy noted that the scales were developed through a bottom-up empirical approach rather than through a top-down theoretical one. This is an important distinction, but we disagree that a top-down approach would have been preferable. The AP/HP scales are direct operationalizations of two empirically derived facets, which were identified using a variety of methods, including (a) generating a set of pride-related feeling states by asking lay people to list words relevant to pride and identify the emotion experienced by individuals expressing pride nonverbally, then obtaining pride-prototypicality ratings for these words; (b) analyses of how these pride-related words cluster together (based on semantic similarity ratings); and (c) analyses of the feelings individuals report when recalling autobiographical memories of pride experiences (Tracy & Robins, 2007).

More specifically, five of the seven items on the HP scale (i.e., arrogant, conceited, stuck-up, pompous, egotistical) were generated by participants asked to freely label the pride nonverbal expression (see Tracy & Robins, 2004b, Study 2); three of these were also spontaneously reported by participants asked to list words that describe how they feel when they feel pride, as were the remaining two on the scale (i.e., snobbish, smug; Tracy & Robins, 2007, Study 2). All of these words were then rated by a team of judges (from the same population—undergraduate students) as highly prototypical of pride. For AP, three of the seven items (i.e., achieving, confident, accomplished) were used to label the pride nonverbal expression (Tracy & Robins, 2004b, Study 2) and were also spontaneously mentioned by participants free-listing words that describe their pride feelings, as were the remaining four items on the scale (fulfilled, productive, self-worth, successful; Tracy & Robins, 2007, Study 2). By building scales from words that emerged directly from participants’ reports of pride-like words or inferences about others’ pride experiences, we ensured that the AP/HP measure would capture pride as it is defined and understood by lay people.

In all factor and cluster analyses of these words, the same two facets emerged, and we theorized that these two empirically derived facets correspond to the theoretical constructs of AP and HP, based on initial evidence of construct validity (i.e., theoretically predicted external correlates; Tracy & Robins, 2007). Since the scales’ publication, studies from our labs and others have provided further support for the scales’ construct validity with regard to Component 2 of the model (i.e., that one facet is adaptive and socially desirable, and the other is maladaptive and socially undesirable; e.g., Ashton-James & Tracy, 2012; Bureau et al., 2013; Carver & Johnson, 2010; Cheng et al., 2010; Damian & Robins, 2012, 2013; Lange & Crusius, 2015; Mercadante & Tracy, 2022; Sanders et al., 2012; Shi et al., 2015; Tracy et al., 2009; Weidman et al., 2016; Weidman & Kross, 2020; Witkower et al., 2022; see Dickens & Robins, 2022 for meta-analysis results) and with regard to some but not all aspects of Component 3 (i.e., their distinct associations with self-esteem, narcissism, and related self-evaluative processes, their cross-cultural generalizability, and their likely evolutionary function in terms of subserving the attainment of prestige and dominance; Carver & Johnson, 2010; Cheng et al., 2010; Shi et al., 2015; Tracy et al., 2009; Witkower et al., 2022). Although these studies do not provide definitive validation evidence, they do suggest that the scales meet several generally accepted criteria for establishing validity.

Furthermore, several studies provide evidence for the scales’ construct validity by manipulating AP and HP with prompts based on their distinct conceptualizations or by experimentally testing predictions that emerge from the theory. For example, Ashton-James and Tracy (2012) manipulated AP and HP using a relived emotion task that emphasized the attributional distinction; the AP manipulation asked participants to write about a time when “you were doing really well in your courses as a result of your efforts” and “you just felt good about your accomplishments in this course—but you did not feel superior to any of your classmates,” and the HP manipulation referred to a time when “you were doing really well in your courses, and finding you didn’t even have to work hard—you just felt naturally talented” and “even superior to many of your classmates” (p. 471). As expected, the HP manipulation
led to significantly higher HP and lower AP compared to the AP manipulation, which led to the opposite pattern. Furthermore, Mercadante and Tracy (2022) tested the hypothesis that hubristically proud individuals use strategic dishonesty to acquire social rank in response to status threats—a suggestion based on the original two-facet theory’s conceptualization of HP. Participants who scored high on trait HP tended to cheat to gain status when their own status was threatened, and this effect held controlling for narcissism, Machiavellianism, and psychopathy, suggesting that this behavioral response is specific to HP.

Research to date using the AP/HP scales thus supports many but not all the tenets of the original theoretical model, which in turn supports the validity of the scales as measures of the theorized constructs. As a result, we believe that Dickens and Murphy overstated the limitations of the extant scales and drew conclusions that are problematic for the body of empirical research that has emerged using them. Perhaps most notably, in their discussion of our bottom-up approach to developing the scales, Dickens and Murphy confute constructs developed in this manner with folk constructs. Not all constructs that are empirically uncovered from participants’ self-reports of a psychological phenomenon are necessarily folk constructs. In fact, uncovering folk constructs of AP/HP would require a different process, in which participants report their understanding of the terms “authentic pride” and “hubristic pride” and measurement tools are developed based on those responses. In contrast, the bottom-up procedure we used asked participants to report words related to their own and others’ pride experiences, then analyzed the semantic similarity of these words and the clustering of participants’ ratings of each word during their own pride experiences. This approach resulted in scales that are based on actual experiences and not folk beliefs about what those terms mean, and which were uncovered empirically rather than mandated a priori, as required by a top-down approach.

Dickens and Murphy raised other concerns with our bottom-up approach that apply to any measure developed using these procedures. We appreciate their point about the cultural boundedness of word usage among certain groups (see Fiske, 2020), and it is undeniable that the populations sampled by Tracy and Robins (2007; i.e., American undergraduates) lack cultural diversity. Nonetheless, it is encouraging that the two-factor structure of pride has been replicated in several languages besides English and countries, including China, Korea, Germany, and Poland (Körner & Schütz, 2021; Shi et al., 2015; Słaskiet et al., 2021). Several of these studies used combined emic/etic approaches, to avoid simply importing American versions of the constructs to cultures that might understand and experience pride differently.

We also appreciate Dickens and Murphy’s point regarding the polysemous nature of common emotion terms, and we agree that these terms can be used in different ways to communicate different experiences and feelings across contexts. However, we take confidence in the fact that many participants across multiple studies using different procedures, measurement approaches, and analysis techniques have produced the same results (Tracy & Robins, 2007). The commonalities that emerged across these endeavors speak to a broad consensus in the words that capture the core of these constructs. Of course, it is possible that context-dependent measures of AP/HP would be more useful for certain investigations, but the empirical derivation of the extant scales, combined with the accumulated construct validity evidence that has emerged over the past 15 years, supports the validity and utility of these instruments.

Notably, Dickens and Murphy propose an alternative way to construct AP/HP scales, using a top-down approach that incorporates the theorized elicitation process into the scale items. Although top-down approaches are useful for ensuring that resulting scales accurately represent theorized constructs, they also can produce measures that necessarily confirm the theory. In the context of AP/HP, items should tap into the broader psychological meaning of the constructs—an adaptive pride facet rooted in actual accomplishments and a maladaptive pride facet rooted in narcissistic self-aggrandizement—but specific eliciting processes (e.g., effort vs. ability attributions) should not be part of the items’ content. Scales constructed using the kinds of items Dickens and Murphy suggest (e.g., “I feel very accomplished thanks to the hard work I put into my projects”; “I feel fulfilled because my natural giftedness contributes to my success”) would conflate the measure with the theory, precluding any independent test of the theory. In contrast, our bottom-up approach provides a means of evaluating (and invalidating) our theoretical model; indeed, as noted previously, studies using these scales have shown that effort vs. ability attributions do not reliably distinguish between the facets (e.g., Dickens & Robins, 2022; Holbrook et al., 2014).

Dickens and Murphy further argued that the HP scale, in particular, is problematic because it does not measure responses that occur only to success. However, this is true of most emotion scales. Emotions are typically defined as coordinated sets of thoughts, feelings, behavior, and physiology that are elicited by a specific appraisal process, but not by a specific situation (e.g., Nesse, 2014; Roseman & Smith, 2001). There is a prototypical situational trigger for fear (i.e., danger) but fear can occur in situations where no danger exists. In the same way, one can feel like they are achieving even when they are not. Self-report emotion scales should therefore avoid measuring specific eliciting situations, and instead capture a particular set of subjective feelings that can occur across different situations.

Dickens and Murphy also suggested that “the HP scale is not a pride scale at all”. This conclusion, however, necessarily depends on what one considers to be within the domain of pride. Clearly, lay people view the scale items as representative of pride, given that they provided the list of words related to pride that now constitute the scale items. Although Dickens and Murphy suggest that a valid HP scale should base its items on researcher-derived definitions, our use of feelings reported by lay people—the presumable targets of the scale—does not invalidate the scale. For certain psychological constructs researcher-derived content is likely to be preferable, but it is not clear why this should be the case for HP, a feeling state, defined by lay people and the model’s originators as easily understood concepts of arrogance, conceit, and egotism.

Dickens and Murphy’s other major criticism is that “the HP scale items do not assess such feelings themselves: Instead, they reflect highly pejorative self-evaluations that one behaves in a socially distasteful manner” (p. 10). This statement may well be true, but no evidence exists to support it. It is equally likely to be true that the HP items measure feelings themselves. Ever since Holbrook et al. (2014) offered this suggestion without testing it, we have been eager to see innovative studies seeking to empirically determine whether individuals who endorse the HP items do so because they in fact feel that way or because they use items like
“arrogant” and “egotistical” to indicate their own negative self-evaluation (presumably for feeling arrogant and egotistical). Dickens and Murphy repeat Holbrook et al.’s (2014) concern without conducting such studies, even though this issue applies to every self-report measure of emotion, as we noted in our response to Holbrook et al. (see Tracy & Robins, 2014). Without new studies probing into participants’ cognitive processes when they complete the scale (a research direction Dickens and Murphy suggested), all we can know is that the HP scale items came from participants’ reports of actual pride experiences, and, since the scale was developed, thousands of participants in multiple countries have demonstrated an ability to reliably rate the extent to which those items describe their current and trait-like feelings.

In addition to these major concerns, Dickens and Murphy raise several other issues with the HP scale. First, they argued that the scale cannot measure pride because it is positively correlated with negative affect. Yet it is not clear why positive valence is a necessary criterion for pride, especially when emotions are measured as trait-like tendencies that occur across situations, rather than as momentary state-like experiences. In many cultures pride is seen as a problematic feeling state, which could easily lead to unpleasant feelings (perhaps alongside pleasant ones). Furthermore, HP might be pleasurable in the short term but dysfunctional in the long term; such a pattern would result in positive correlations with tendencies to feel negative affect at the between-person level. Note that positive correlations with negative affect do not necessarily indicate the simultaneous experience of HP and negative affect, given that these correlations are typically conducted at the between-, not within-person level. Notably, Tracy and Robins (2007) showed that the empirical clustering of AP and HP as distinct dimensions holds even after controlling for valence, so neither facet can be characterized as simply positive or negative valence factors.

Second, Dickens and Murphy argued that the HP scale lacks convergent validity because it does not correlate highly with the (very few) other existing pride scales. Yet those scales were not developed to assess HP or the distinction between HP and AP; they were developed to measure “pride”, which, prior to our model, was exclusively defined very similarly to our model’s definition of AP. These scales therefore cannot be used to evaluate convergent validity with HP, because they do not measure the same construct. If every measure of a newly conceptualized construct was required to converge with measures of prior conceptualizations of the construct, we would make little progress in the development and assessment of new constructs.

Third, Dickens and Murphy critiqued the HP scale for being highly skewed, yet this is typical of scales assessing socially undesirable constructs (as well as scales assessing highly socially desirable constructs). The critical issue is not the degree of skewness, but the amount of variance. The standard deviation of the HP scale is well within the range of most individual-difference measures in psychology, albeit somewhat lower than average. Moreover, the true population distribution of HP probably is highly skewed, because most people are not very narcissistic, and thus not prone to hubristic pride. Based on the test information function presented by Dickens and Murphy, the HP scale does not do a particularly good job of discriminating among people who are very low vs. low in hubristic pride. However, it reliably discriminates among the rest of the distribution: people who are below average vs. moderate vs. high in hubristic pride—which is what we would expect for a scale assessing a psychological tendency that is not particularly common in the population. As for the concern that “all the scale items are highly pejorative” (p. 14), this is necessarily the case when assessing a socially undesirable construct.

Finally, Dickens and Murphy’s speculations about various potential sources of variation in HP scores is informative, but this interpretative analysis could be performed for almost any self-report scale. For example, diverse processes can lead to high scores on a self-esteem scale, but this does not mean that the measure does not assess a coherent construct. Fortunately, we have empirical methods for examining the coherence of a measure, so there is no need to speculate about this issue; in this context, it is worth noting that the HP scale is necessarily an empirically coherent measure, because empirical coherence was the primary criterion used to develop the scale (via factor analyses, cluster analyses, alpha reliability analyses; see Tracy & Robins, 2007).

Nonetheless, despite our concerns with Dickens and Murphy’s critique, we agree that the AP/HP scales may not be optimal for assessing the theorized constructs of AP/HP, largely because of the (valid) questions that have been raised regarding what the HP scale is measuring—whether it is a distinctive feeling state or a negative evaluation of oneself for feeling that state. At the same time, we are doubtful whether it would be possible to develop a self-report measure of HP that is not: (a) positively skewed, (b) lacking in convergent validity with prior measures of pride that did not aim to capture the emotion’s hubristic form, (c) negatively valenced, (d) producing scores driven by numerous potential sources of variation, and (e) unambiguously measuring an actual and distinctive emotional experience and not a self-evaluation. Nonetheless, this is an empirical question for future research.

On this note, Gerasimova (2022) provided a template for how such work might proceed. To evaluate the validity of measurement tools, Gerasimova (2022) suggested creating a living document of validity evidence that exists alongside a given measure. The document might begin with a set of claims about the appropriate interpretations and uses of the measure, along with descriptions of evidence required to support those claims. As evidence accumulates, the document is updated accordingly. For example, one might claim that a valid HP scale should show positive correlations with antisocial behaviors; currently, evidence in support of this claim could be drawn from several sources (e.g., Mercadante & Tracy, 2022; Tracy et al., 2009). Ideally, tests of a priori claims like these would be preregistered, though it is too late to do so for most extant research validating the AP/HP scales (but see Mercadante & Tracy, 2022; which reports four preregistered studies). When the entirety of accumulated evidence can be easily seen and evaluated in aggregate, problem areas reveal themselves (as has occurred for claims regarding AP/HP’s distinct attributional antecedents), and new preregistered validation studies can be conducted to address these issues. Ultimately, this process should lead to needed adjustments to claims about the appropriate interpretations and uses of the HP scale, or to the measure itself. In our view, this approach would be a productive way for researchers to move forward, given the issues articulated by Dickens and Murphy. We look forward to future research in this vein that might ultimately improve the existing scales or develop new ones.
References


Received August 20, 2022
Accepted September 12, 2022